

مركز البحوث والدراسات

تطوير المقاييس
النظرية والتطبيق

تأليف:
روبرت ف. ديفيليس

ترجمة
د. سعد بن سعيد القحطاني
أ.د. طارق عطية عبدالرحمن

راجع الترجمة
أ.د. عبدالرحمن بن سليمان العنقري

بسم الله الرحمن الرحيم



مركز البحوث والدراسات

تطوير المقاييس

النظرية والتطبيق

تأليف:
روبرت ف. ديفيليس

ترجمة
د. سعد بن سعيد القحطاني
أ.د. طارق عطية عبدالرحمن

راجع الترجمة
أ.د. عبدالرحمن بن سليمان العنقري

١٤٤٢هـ - ٢٠٢١م

بطاقة الفهرسة


معهد الإدارة العامة، ١٤٤٢هـ.
فهرسة مكتبة الملك فهد الوطنية أثناء النشر.
ديفيليس، روبرت ف
تطوير المقاييس: النظرية والتطبيق. / روبرت ف ديفيليس؛ سعد بن سعيد القحطاني؛ طارق عطيـة عبدالرحمـن؛ عبدالرحمـن سـليمان العنقري. - الرياض، ١٤٤٢هـ.
٣٤٤ ص؛ ١٧ سم × ٢٤ سم.
ردمك: ٧ - ٢٩ - ٨٢٧٦ - ٦٠٣ - ٩٧٨
١- الموازيـن والمقاييس أ. القحطاني، سـعد بن سـعيد (مترجم) ب. عبدالرحمن، طارق عطية (مترجـم) ج. العنقري، عبدالرحمن سـليمان (مراجع) د. العنوان
ديوي ٣٨٩,١ ١٤٤٢/٣٠٨٥

رقم الإيداع: ١٤٤٢/٣٠٨٥
ردمـــــك: ٧ - ٢٩ - ٨٢٧٦ - ٦٠٣ - ٩٧٨

هذه ترجمة لكتاب:

SCALE DEVELOPMENT

Theory and Applications

By

Robert F. DeVellis



قائمة المحتويات المختصرة

الموضوعات	الصفحة



٦

قائمة المحتويات التفصيلية

الموضوعات	الصفحة

الموضوعات	الصفحة

الموضوعات	الصفحة

الموضوعات	الصفحة

الموضوعات	الصفحة

الموضوعات	الصفحة

الموضوعات	الصفحة

مقدمة:

تطويـر المقاييس مجال يتطـور بسرعة مع ظهور مناهج جديدة واكتسـاب المزيد من القوة. وقد نشـأت هذه التغييرات من مجموعة متنوعـة من المصادر، بما في ذلك الجهود الكبـيرة لتوفير المزيد مـن أدوات القياس الموحدة بشـكل أكبر عبر مجموعة واسـعة من السـياقات. فمن ناحية، تتسـم هذه التغييرات بالأهمية والإثارة. ومن ناحية أخرى، فإنها تجعـل مراجعـة وتنقيح هذا الكتاب تحدياً إلى حد كبير. وكما هو الحال مع الإصدارات أو الطبعات السابقة، فإن هدفي هو تقديم معلومات دقيقة أو معقدة بطريقة يمكن لمجموعة مـن الباحثين الوصول إليها. وفي الوقت نفسـه، ومع كل مراجعـة وتنقيح، حاولت جاهداً تحديث المادة العلمية لتعكس الاتجاهات الحالية في هذا المجال. ويتعارض هذان الهدفان في بعـض الأحيـان تعارضاً تاماً، حيـث إن بعض التطورات الجديـدة تنطوي على مفاهيم ومنهجيات متقدمة ومتخصصة نوعاً ما. وقد سعيت إلى حل هذه المعضلة عن طريق توفير سـياق للتطورات الجديدة، وعن طريق التمييز بشكل واضح بين أنشطة القياس المحدودة والضيقة وأنشـطة القياس الأوسع نطاقًا والنُهج والمداخل التقنية التي يمكن أن تناسب كلاً منها. وبطبيعة الحال، فقد اقتصرت مناقشتي لهذه القضايا والمسائل على سياق نطاق الكتاب وهدفـه الذي يعد بمثابة نظرة عامة على النُّهج السـيكومترية بدلاً من المراجعة الشـاملة.

وكانـت المحصلـة النهائية هي أن كل فصل يحتوي على مـادة جديدة مضافة إليه. في الفصل الأول، نظرة عامة، يتضمن مراجعتي واستعراضي لتاريخ القياس الآن ملخصًا لكيفية تصنيف الأمراض العقلية من العصور القديمة إلى الوقت الحاضر، مع التركيز على الأنشـطة الأكـثر حداثة. ويكشـف هـذا النقاش عن الفـرص الضائعة وغيرها من الفـرص التي تم استغلالها.

في الفصل الثاني، فهم المتغير الكامن، قمت بإضافة قسمٍ يتناول بشكل مباشر أكبر الآثار المترتبة على اختيار النموذج السـببي الذي يربط البنود بمتغيراتها الكامنة. وقد لاحظت أن هذا الاعتبار المهم والواضح على ما يبدو، كثيراً ما يتم تجاهله وإغفاله. وبعض النماذج التي تربط المتغيرات بمؤشراتها تكون أكثر ملاءمةً لنُهج القياس المباشرة أكثر من غيرها. وبناءً على ذلك، أعتقد أن هذه النقطة تستدعي وتحتاج مزيداً من التركيز.

وتمثـل التغييرات في الفصل الثالـث، الثبات، التطور السريع في التفكير في بعض مجالات القياس. وقد ظهرت في السنوات الأخيرة مجموعة كبيرة من الأعمال التي تشكك في استخدام معامـل «كرونباخ ألفا». وعلى الرغم من أنني قـد أشرت إلى هذه المنظورات الجديدة في الطبعة السابقة، إلا أنني قد شعرت بالحاجة إلى التوسع في تلك المناقشة في الطبعة الحالية مـن الكتاب. وهذا يتضمن تسـليط الضوء على بعض الانتقـادات والظروف التي تنطبق عليهـا، فضلاً عن مناقشـة أكثر تفصيلاً لبعض بدائل معامل «ألفـا»، مثل معامل»أوميغا»، والذي اكتسب قدراً أكبر من الاعتراف. وقد حاولت أن أحافظ على أن تكون هذه المناقشة في المقام الأول مفاهيمية ونظرية وليست رياضية، وبدلاً من تكريس أو تخصيص الصفحات للعمليات الحسابية التفصيلية، فقد وجَّهت القراء إلى المصادر التي ستساعدهم في استخدام أدوات مثل معامل «أوميغا» إذا اختاروا ذلك. لقد قمت أيضاً بإضافة مناقشـة حول مدى ثبـات درجات التغيير المحسـوبة كفرق بين درجتين من مقيـاس واحد تم تطبيقه أكثر من مرة. وألخص من الناحية النظرية ماهية ثبات درجة التغيير، وأسـلط الضوء على الظروف عندما يكون في استخدامها مشكلة، وأقترح أساليب تحليلية توفر بعض الحماية ضد الثبات المخفف الناشئ عن استخدام درجات التغيير الأولية أو الخام.

في الفصل الرابع، الصدق، قمت بتوسيع نطاق معالجتي لصدق المحتوى بشكل كبير. وأؤكد عـلى أهمية الحفاظ على علاقة وثيقة بين المتغير محل الاهتمام والمقياس المسـتخدم لقياسـه كمياً. كما أناقش أيضاً المخاطر المحتملة لوضع مفهوم للمتغير على نطاق أوسع من اللازم.

وفي الفصل الخامس، القواعد الإرشادية لتطوير المقياس، قدمت عدداً قليلاً من المسائل إما مزيداً من التركيز أو قدمت مزيداً من التفاصيل إلى حد ما. وحاولت أيضاً ربط المسائل العملية التي يركز عليها هذا الفصل ببعض المسـائل المفاهيمية «النظرية» المثارة في فصول أخرى.

في الفصل السـادس، التحليل العاملي، واجهت مرة أخرى معضلة الموازنة بين الأنشـطة الجديـدة المهمـة في هذا المجال وبين الحفاظ على الأمور لتكون في متناول القارئ وسـهلة القراءة. وفي النهاية، اخترت إضافة قسم جديد حول نموذج ثنائي العوامل bifactor. ولقد فعلـت ذلك لأنه فيما يبدو أصبح يسـتخدم على نطاق أوسـع لتقييـم المتغيرات المعقدة أو المركبة التي تشـتمل عـلى جوانب أكثر عمومية وأكثر تحديـداً. ولا تهدف تغطيتي إلى

جعــل القراء محللين مقتدرين، بل فهم ســبب مواجهتهم لهذا النمــوذج (ونموذج العامل الهرمــي المماثل من الناحية المفاهيمية) في قراءتهم. وعلى وجه الخصوص في النُهج القائمة على نظرية اســتجابة البند، يمكن للنماذج ثنائية العوامل أن تخدم غرضاً محدداً (المساعدة على تحقيق أحادية البعد) والتي شعرت أنه يستحق بعض التوضيح والتفسير.

يقدم الفصل الســابع مرة أخرى نظرة عامة على نظرية استجابة البند. وكما فعلت مع الفصل الســادس، واجهت قرارًا بشــأن ما يجب تضمينه ومعالجته، وكنت أرغب في إضافة المزيد من التفاصيل مع الحفاظ على الأســلوب والمســار العام للكتاب. لقد اخترت تقديم مناقشــة جديدة حول ثيتا theta، أي قوة الســمة التي يجري تقييمهــا، ودورها في اختبار التكيف المحوسب computerized adaptive testing (CAT). ولم يصبح اختبار التكيف المحوســب CAT أكثر انتشــارًا فحســب، بل إنه، في بعض النواحي، يمثل وســيلة مختصرة لتلخيص بعض الاختلافات بين نظرية استجابة البندIRT والنُهج الكلاسيكية للقياس.

يناقش الفصل الثامن مرة أخرى القياس في الســياق الأوسع. وكما فكرت في العديد من التغييرات في نهج القياس التي طرأت على أساليب القياس منذ ظهور الطبعة الأولى من هذا الكتاب، فقد أدهشتني حقيقة أن العديد منها قد جاء نتيجة لجهود القياس واسعة النطاق large-scale measurement efforts. هــذه الجهود جديرة بالاهتمام ومهمة ولكنها على الأرجح ليست نوع القياس الذي سيشارك فيه معظم الباحثين. ولذلك، شعرت أنه سيكون مــن المفيد مقارنة «القيــاس الصغير» مع «القياس الكبير». وأؤكــد في هذا الصدد على أن كليهما قادر على توفير معلومات ذات مغزى وأن لكل منهما مزاياه وعيوبه. كما أؤكد على أن إبقاء هذه الاختلافات في الاعتبار منذ بداية مشروع القياس ســيزيد في كثير من الأحيان من احتمال النجاح ويمكن أيضاً أن يجعل المهمة أسهل بكثير بطرق مختلفة.

كما قمت بإضافة تمارين إضافية في نهاية الفصول. وآمل أن تساعد هذه التمارين القراء على تقييم فهمهم للمواد، وأن تحفزهم أيضاً على مناقشة القضايا المهمة.

لقــد احتفظت بالميزات التي وجدها القراء مفيدة للغاية في الطبعات الســابقة. وهذا يعني أن التركيز يظل في المقام الأول على المفاهيم بدلاً من العمليات الحســابية. وكما هو الحال في الطبعات السابقة، فإن المساحة المحدودة التي أكرسها للصيغ والمعادلات ينبغي أن

تساعد القراء على رؤية ما هو أبعد من الحروف والرموز اليونانية لفهم المفاهيم التي تعبر عنها تلك الصيغ. وكلما كان ذلك ممكناً، أستخدم التشبيهات لشرح العلاقات بين المفاهيم، وهو النهج الأسلوبي الذي أخبرني القراء السابقون أنه مفيد. لذلك؛ ففي حين تمت إضافة المزيد من المواد «الموضوعات» المتقدمة من أجل إبقاء القراء على اطلاع على الاتجاهات الجديدة في القياس، فقد بذلت قصارى جهدي للحفاظ على إمكانية الوصول إلى الطبعات السابقة وسهولة قراءتها.

شكر وتقدير:

ترغب شركة SAGE Publishing في الاعتراف بالمساهمات القيمة التي قدمها المراجعون التالية أسماؤهم: جوري كولبرت – جيرتز Jorie Colbert-Gertz، كلية جونز هوبكنز للتعليم؛ وستيفن دبليو ديتمور Stephen W. Dittmore، جامعة أركنساس؛ غاري دي إيليس Gary D. Ellis، جامعة تكساس إيه آند إم؛ جوستين جروس Justin H. Gross، جامعة ماساتشوستس أمهيرست؛ بنيامين هيل Benjamin D. Hill، جامعة جنوب ألاباما؛ وجوزيف كوش Joseph C. Kush، جامعة دوكين.

نبذة عن المؤلف:

قبل تقاعده في عام ٢٠١٢، عمل روبرت ف. ديفيليس أستاذاً في قسم السلوك الصحي (كلية جيلينجز للصحة العامة العالمية Gillings School of Global Public Health) في جامعة نورث كارولينا في تشابل هيل. يتمتع الدكتور ديفيليس بخبرة تزيد عن ٣٥ عاماً في قياس المتغيرات النفسية والاجتماعية، وقد كان عضواً نشطاً في اتحاد نظام معلومات قياس نتائج تقارير المرضى (PROMIS)، وهو عبارة عن مبادرة لخارطة طريق للمعاهد الوطنية للصحة (NIH) متعددة المواقع تهدف إلى تحديد وتعديل واختبار ونشر المقاييس النهائية لاستخدامها من قبل باحثي المعاهد الوطنية للصحة، وكان دوره في اتحاد نظام معلومات قياس نتائج تقارير المرضى كرئيس مجال الشبكة الواسعة للنتائج والمخرجات الاجتماعية. كما عمل في مجلس إدارة شعبة علم النفس الصحي التابع للجمعية الأمريكية لعلم النفس (٣٨)، وفي قسم دراسات البحوث السريرية/النتائج/العلاجية التابع لمؤسسة

التهـاب المفاصـل "الروماتيـزم"، وفي المجلس الاستشـاري لمبادرة التميز في قياس شـؤون المحاربين القدامى، وقد حصل على جائزة العالم المتميز لعام ٢٠٠٥ من جمعية المتخصصين في أمراض الروماتيزم، وهو محرر مشارك لرعاية وبحوث التهاب المفاصل. بالإضافة إلى ذلك، عمل كمحرر زائر، أو محرر مشارك زائر، أو مراجع لأكثر من عشرين مجلة أخرى. وقد عمل كباحث رئيسي أو باحث مشـارك منذ أوائل الثمانينيات في سلسلة من المشروعات البحثية التي تمولها الحكومة الاتحادية والمؤسسـات الخاصة، وما زال يعمل بشـكل نشط في دوره التحريري وكمستشار في مجموعة متنوعة من المشاريع.



الفصل الأول
نظرة عامة

يُعـد القياس من الأمور الحيوية عبر طائفة واسـعة من سـياقات البحوث الاجتماعية. على سبيل المثال، فكّر في الحالات الافتراضية التالية:

١- تواجـه المتخصصة في علم النفس الطبي معضلة شـائعة: مقياس القياس الذي تحتاجه غير موجود على ما يبدو. وتتطلب دراسـتها أن يكون لديها مقياس يسـتطيع أن يفرق ويميـز بين ما يريد الأفراد حدوثـه وما يتوقعون أن يحدث عندما يـرون طبيباً. وتبين أبحاثها أن الدراسـات السـابقة اسـتخدمت المقاييس التي أربكت وأثارت بدون قصد هاتـين الفكرتين. ولا يبـدو أن المقاييس الموجودة تقوم بهذا التمييز على وجه التحديد بالطريقة التي تريدها المتخصصة في علم النفس الطبي. وعلى الرغم من أنها تسـتطيع تكوين وإنشـاء بعض الأسئلة التي يبدو إمكانية الاستفادة منها في التمييز بين ما يريده الفرد وما يتوقعه، فإنها تخشى من أن البنود أو الأسـئلة «التي تم تكوينها» قد لا تكون مؤشرات موثوقة وصالحة لهذه المفاهيم.

٢- أن المتخصـص في علـم الأوبئة غير متأكد من كيفية المضي قدمـاً والمتابعة. وهو يجري تحليلات ثانوية بشـأن مجموعة كبيرة من البيانات تستند إلى دراسة استقصائية وطنيه بشأن الصحة. وقال إنه يود أن يفحص العلاقة بين جوانب معينة من الإجهاد (الضغط) النفسي المـدرك والحالة الصحية. على الرغم من عدم تضمين مجموعه من البنود التي تم تصميمها كمقياس للإجهاد أو الضغط النفسي في الدراسة الاستقصائية الأصلية، إلا أن العديـد من البنود التي كانت تهدف في الأصل إلى قياس متغيرات أخرى يبدو أنه يمكن الاسـتفادة من محتواها المرتبط بالضغـط النفسي. وقد يكون من الممكن تجميع هذه البنود في مقياس يتمتع بالثبات والصدق لقياس الإجهاد أو الضغط النفسي. ومع ذلك، إذا كانت البنود المجمعة تشـكل مقياسـاً ضعيفاً للإجهاد النفسي، فقد يتوصل الباحث إلى استنتاجات خاطئة.

٣- يشعر فريق التسويق بالإحباط في محاولاته لتخطيط حملة لخط أو مجموعة جديدة من ألعاب الأطفال الرضع عالية السعر. وقد أشارت مجموعات التركيز (focus groups) إلى أن قرارات الشراء التي يتخذها الآباء تتأثر بشـــده بالأهمية التعليمية الظاهرة للألعاب مـــن هذا النوع. ويعتقد فريق التســويق أن الآباء الذين لديهـــم طموحات وتطلعات تعليميـــة ومهنية عالية لأطفالهم الرضع ســيكونون أكثر اهتمامـــاً وانجذاباً لهذا الخط الجديـــد من الألعاب. ولذلك، يرغب فريق التســـويق في تقييـــم هذه التطلعات ضمن عينـــه كبـــيره ومنتشرة جغرافيًّا من الآبـــاء. وتعتبر مجموعات التركيـــز الإضافية مرهقة للغاية للوصول إلى عينة كبيرة بما فيه الكفاية من المستهلكين.

في كل حالـــة من هـــذه الحالات، يواجه الأشـــخاص المهتمون ببعض المجـــالات العلمية مشكلة تتعلق بالقياس. ولا يهتم أي من هؤلاء الباحثين في المقام الأول بالقياس في حد ذاته. ومع ذلك، يجب على كل منهم أن يجد وســـيلة لقياس ظاهره معينة قبل التصدي لتحقيق هـــدف البحـــث الرئيسي. وفي كل حالة، تكون أدوات القياس «الجاهزة أو المتوفرة» إما غير مناسبة أو غير متاحة، ويدرك جميع الباحثين أن تبني مناهج القياس العشوائية ينطوي على مخاطـــر إنتاج بيانات غير دقيقة. وبالتالي، فإن تطوير أدوات القياس الخاصة بهم يبدو أنه هو الخيار الوحيد المتبقي أمامهم.

وقـــد واجه العديـــد من الباحثين في العلوم الاجتماعية مشـــاكل مماثلة. وتتمثل إحدى الاســـتجابات الشـــائعة للغاية لهذه الأنواع من المشـــاكل في الاعتماد عـــلى الأدوات القائمة المشـــكوك في ملاءمتها. وهناك طريقة أخرى للاســـتجابة لهذه المشاكل تتمثل في افتراض أن بنود الاســـتبيان المطورة حديثًا «التي تبدو ســـليمة وصحيحة» ستقوم بعمل قياس مناسب. ومن الأعذار الشائعة لضعف إستراتيجيات القياس هي عدم الارتياح أو عدم الإلمام بأساليب تطويـــر أدوات قياس تتمتع بالثبات والصدق وعدم إمكانية الوصول إلى المعلومات العملية حـــول هذا الموضـــوع. وقد تؤدي محـــاولات الباحثين الرامية إلى اكتســـاب مهارات تطوير القياس إلى توجيههم إما إلى مصادر غامضة مخصصة أساســـاً للمتخصصين في القياس أو إلى المعلومات التي تكون عامة جداً بحيث لا تكون مفيدة، وتم إعداد هذا الكتاب ليكون بديلاً لتلك الخيارات.

المنظورات العامة بشأن القياس:

يُعد القياس نشــاطاً أساســياً للعلوم. ونحصل على المعرفة حول الأشــخاص، والأشــياء، والأحــداث، والعمليــات من خــلال ملاحظتها ورصدهــا، ويتطلب فهم هــذه الملاحظات باستمرار قياسها بصورة كمية (بمعنى أننا نقيس الأشياء التي نهتم بها من الناحية العلمية). وتتفاعل عملية القياس والأســئلة العلمية الأوســع نطاقاً التــي تخدمها مع بعضها البعض، وغالبــاً مــا تكون الحدود بينها غــير ملحوظة، ويحدث هذا، على ســبيل المثال، عندما يتم اكتشاف كيان جديد أو تنقيحه في سياق القياس أو عندما يفسر المنطق الذي ينطوي عليه تحديد كيفية القياس الكمي لظاهرة موضع الاهتمام في ضوء الظاهرة نفســها. فمثلاً، قام سميث، وإيرب، وديفيليس (Smith, Earp, and DeVellis, 1995) ببحث تصورات النساء بشــأن الضرب. وقد اقترح نموذج تصوري (مفاهيمي) مســبق قائم عــلى التحليل النظري ســتة مكونات متميزة لهذه التصورات. وقد أشار العمل التجريبي والذي استهدف تطوير مقياس لقياس هذه التصورات إلى أنه - بين النســاء اللاتي تعرضن للضرب والنســاء اللاتي لم يتعرضن له - كانت هناك تصورات أبسط بكثير هي السائدة: لقد شرح مفهوم واحد بدقة كيف اســتجابت النساء المشاركات في الدراسة إلى ٣٧ بنداً من أصل ٤٠ بنداً تم استخدامها. وتشــير هذه النتيجة إلى أن ما اعتبره الباحثون مجموعة معقدة من المتغيرات كانت تنظر إليه النساء اللائي يعشن بالمجتمع المحلي في الواقع أنه ظاهرة واحدة أوسع نطاقاً. وهكذا، فإنه في ســياق استنباط وسيلة لقياس تصورات النســاء حول الضرب، اكتشفنا شيئاً جديداً حول بنية تلك التصورات.

ويقــول دانكن (Duncan, 1984): إن جذور القيــاس تكمن في العمليات الاجتماعية، وأن هذه العمليات وقياســها في الواقع تســبق العلم: "كل القياس ... هو قياس اجتماعي. والمقاييــس الماديــة طُوّرت لأغراض اجتماعيــة" (ص ٣٥). وفي إشــارة إلى عمليات القياس الاجتماعي الرســمية المبكرة، مثل التصويت وإجراء التعــداد، ونظم التقدم الوظيفي، ذكر «دانكن» أن أصولها تبدو وكأنها تمثل محاولات لتلبية الاحتياجات الإنسانية اليومية، وليس مجــرد تجارب أجريت لتلبية الفضول العلمي. وقد ذهب إلى القول بأنه يمكن اســتخلاص عمليات مماثلة.

> في تاريــخ الفيزياء: فقد تم قياس الطول أو المســافة والمســاحة والحجم والوزن والوقت من قبل الشــعوب القديمة في ســياق حل المشكلات العملية والاجتماعية، وقد بُنيت العلوم الفيزيائية على أسس تلك الإنجازات (ص ١٠٦).

ومهما كانت الدوافع الأولية، فإن كل مجال من مجالات العلوم يطور مجموعه إجراءات القياس الخاصة به. فقد طور علم الفيزياء، على ســبيل المثال، أساليب ومعدات متخصصة للكشــف عن الجزيئات شبه الذرية. وفي إطار العلوم السلوكية والاجتماعية، تطور القياس النفسي (psychometrics) كتخصص فرعي معني بقياس الظواهر النفســية والاجتماعية. وعادةً ما يكون إجراء القياس المســتخدم هو الاستبانة، وتشكل متغيرات الاهتمام جزءاً من إطار نظري أوسع نطاقاً.

الأصول التاريخية للقياس في العلوم الاجتماعية:

أمثله مبكرة:

المنطــق الســليم والســجل التاريخي يؤيدان ادعــاء دنكان (Duncan) بــأن الضرورة الاجتماعيــة أدت إلى تطويــر القياس قبل ظهور العلم. ولا شــك أن بعض أشــكال القياس كانــت جــزءاً من ذخيرة الأنواع لدينا منذ عصور ما قبــل التاريخ. فالبشر الأوائل لابد أنهم استطاعوا تقدير الأشياء والممتلكات والخصوم (المعارضين) على أساس خصائص مثل الحجم. ويذكر «دنكان» Duncan (1984) إشارات الكتاب المقدس في الاهتمام بالقياس (على سبيل المثــال، "التوازن الزائف هو رجس (مكــروه) عند الرب، ولكن الــوزن العادل هو فرحه"، الأمثال ١١:١). ويلاحظ أن كتابات أرسطو تشير إلى المسؤولين المكلفين بالتحقق من الأوزان والقياســات. وتلاحظ اناستاسي Anastasi (1968) أن أسلوب سقراط socratic المستخدم في اليونــان القديمة ينطوي على البحث عن الفهــم بطريقة يمكن اعتبارها اختباراً للمعرفة. يصــف ب. هــ. ديبــوا P. H. DuBois في مقالة له عــام ١٩٦٤ (أعيد طبعهــا في بارنيت، ١٩٧٦) استخدام اختبار الخدمة المدنية في وقت مبكر يصل إلى ٢٢٠٠ قبل الميلاد في الصين. ويستشــهد رايت Wright (1999) بأمثله أخرى عــلى الأهمية التي أولتها العصور القديمة لعملية القياس الدقيق، بما في ذلك "وزن سبعة أمثال" "weight of seven" الذي استندت

إليها الضرائب الإسلامية في القرن السابع. كما يشير أيضاً إلى أن البعض ربط الثورة الفرنسية، جزئياً، بالفلاحين الذين سئموا من ممارسات القياس غير العادلة.

إن الفكـرة القائلة بـأن القياس يمكن أن ينطـوي على خطأ وأنه يمكـن اتخاذ خطوات معينة للحد من ذلك الخطأ تُعد رؤية أحدث. وقد أشـار «بوتشوالد» Buchwald (2006) في اسـتعراضه لتناقضات القياس وتأثيرها عـلى المعرفة، إلى أنه على الرغم من أنه لا يزال في العشرينيات من عمره في أواخر السـتينيات من القرن السـادس عشر وبداية السـبعينيات من القرن السـادس عشر، كان "إسـحاق نيوتن" على ما يبدو هو أول من استخدم متوسط ملاحظـات متعـددة، وكان قصده إنتـاج مقياس أكثر دقـة عندما أسـفرت ملاحظاته عن الظواهر الفلكية عن قيم متناقضة. ومن المثير للاهتمام أنه لم يوثق اسـتخدام المتوسـطات في تقاريـره الأولية ولكنـه أخفى اعتماده عليها لعقود. وقد يكـون هذا الإخفاء ناجماً عن نقـص وانعدام النزاهة أقل مما نشـأ من مجـرد الفهم المحدود للخطـأ ودوره في القياس. وتعليقاً على اسـتخفاف مماثل لعالم فلك آخر بالنسـبة للملاحظات المتناقضة، يقول «ألدر» (Alder, 2002): إنـه حتى في أواخر القرن الثامن عشر، لم يكن إخفاء التناقض في الملاحظة "شـائعاً فقط، بل كان يُعتبر حقًّا لصاحب العلـم أو الموهبة. وكان يُنظر إلى الخطأ على أنه فشل أخلاقي" (ص ٣٠١). وقد خلص «بوتشوالد» (Buchwald, 2006) إلى ملاحظة مماثلة:

> [طريقة عمل العلماء في القرن السـابع عـشر وأوائل القرن الثامن عشر] لا تعتبر الاختلافات نواتج ثانوية حتمية لعملية القياس نفسها، بل دليلاً على مهارة فاشـلة أو غير كافية. ومن المحتمل أن يكون الخطأ في القياس مختلفًا قليـلاً عن السـلوك الخاطئ من أي نـوع؛ إذ يمكن أن يترتـب عليه عواقب أخلاقية، وينبغي إدارته بطرق مناسبة (صفحة ٥٦٦).

لم يكن علماء الفلك الوحيدين الذين يقومون بعمليات رصد منتظمة للظواهر الطبيعية في أواخر القرن السابع عشر وأوائل القرن الثامن عشر. ففي الستينيات من القرن السادس عـشر، كان "جون جراونت" (John Graunt) يقوم بتجميـع معدلات الولادة والوفاة من سجلات التعميد والدفن في مدينة هامبشاير (Hampshire) بإنجلترا. واستخدم "جراونت" أسـلوب حسـاب المتوسـط (على الرغم من أنه ليس الاسـتخدام الشـائع في هذا الوقت)

لتلخيص النتائــج التــي توصل إليهــا. ووفقًــا لبوتشــوالد Buchwald (2006)، كان دافع Graunt لاستخدام المتوسط هو الحصول على قيمة "حقيقية" آنية. وكانت الفكرة هي أن نسبة المواليد إلى الوفيات تمثل لبعض قوانين الطبيعة، غير أن الأحداث التي لا يمكن التنبؤ بها والتي قد تحدث في أي سنة معينة سوف تخفي تلك الحقيقة الأساسية. إن وجهة النظر هذه للملاحظة كنافذة غير مكتملة في حقائق الطبيعة تشير إلى تطور متزايد في كيفية النظر إلى القيــاس: بالإضافــة إلى القيود التي يفرضها الملاحظ، فإنــه يمكن أيضاً لعوامل أخرى أن تفسد المعلومات التي يتم جمعها تجريبياً، وقد تكشف بعض التعديلات التي أدخلت على هذه القيم على نحو أكثر دقة عن الطبيعة الحقيقية للظاهرة محل الاهتمام.

وعــلى الرغم من هذه الرؤى المبكرة، فإنه بعد قرن من الزمن من اســتخدام "نيوتن" لأول مرة لأسلوب المتوسط قبل أن يدرك العلماء على نطاق أوسع أن جميع القياسات كانت عرضة للخطأ وأن أسلوب المتوسط سوف يقلل من مثل هذا الخطأ (Buchwald, 2006). وفقــاً للفيزيائي والمؤلــف "ليونارد مولودينــو" Leonard Mlodinow (2008)، في أواخر القــرن الثامن عشر وأوائل القرن التاســع عشر، أجبرت التطــورات في علم الفلك والفيزياء العلــماء عــلى الاقتراب من الخطأ العشــوائي بشــكل أكــثر منهجية، مــما أدى إلى ظهور الإحصاء الريــاضي (mathematical statistics). وبحلول عام ١٧٧٧، قارن دانييل برنولي «Daniel Bernoulli» (ابــن شــقيق جاكوب برنولي «Jakob Bernoulli» الأكثر شــهرة) توزيعــات القيم التي تم الحصول عليها من الملاحظات الفلكية على مســار ســهام الرمي (archer's arrows)، وتكتلــت أو تجمعت حول نقطة مركزية تقل بشــكل تدريجي على مســافات أكــبر من ذلك المركز. وعلى الرغــم من أن المعالجة النظريــة التي رافقت تلك الملاحظــة كانــت خاطئة في بعــض النواحي، إلا أنهــا تمثل بداية تحليل رســمي للخطأ في القياس (Mlodinow, 2008). ويجادل "بوتشوالد" Buchwald (2006) بأن أحد العيوب الأساســية في تفســيرات القــرن الثامن عشر للخطــأ في القياس هو الفشــل في التمييز بين الخطأ العشوائي والخطأ المنتظم. وحتى بزوغ فجر القرن التالي سوف يظهر فهمٌ أكثر دقة للعشــوائية. ومع هذا الفهم المتزايد للعشوائية جاء التقدم في القياس؛ ومع تقدم القياس، تقدم العلم أيضاً.

ظهور الأساليب الإحصائية ودور الاختبارات العقلية:

يدعـم منظـور «نونـالي» Nunnally (1978) الـرأي القائـل بأن الفهم الأكـثر تطوراً للعشـوائية والاحتـمالات والإحصاء كان ضروريًـا لازدهار القياس، ويقـول إنه على الرغم مـن أن الملاحظات المنتظمة ربما كانت مسـتمرة، فإن عدم وجود طرق إحصائية رسـمية بدرجـة أكبر قد أعاق تطوير علم لقياس القدرات البشريـة حتى النصف الأخير من القرن التاسـع عشر. وقد بدأ تطوير الأساليب الإحصائية المناسبة في القرن التاسع عشر في النهاية من خـلال مشروع "داروين" Darwin بشـأن التطور وملاحظته وقيـاس التباين المنتظم عبر الأنواع. وقام ابن عم داروين، السـير "فرانسـيس غالتون" Francis Galton، بتوسيع الملاحظة المنسقة للاختلافات إلى البشر. وكان اهتمام "غالتون" الرئيسي هو وراثة الصفات التشريحيـة والفكريـة. وكان "كارل بيرسـون" Karl Pearson، الـذي يعتـبره الكثـيرون "مؤسـس الإحصاء" (على سبيل المثال، ألين وين، Allen& Yen, 1979, p. 3)، زميلًا صغيرًا لجالتون. وطور بيرسون الأدوات الرياضية - بما في ذلك معامل ارتباط حاصل ضرب العزوم (Product-Moment Correlation Coefficient) والذي يحمل اسمه، أي معامل ارتباط بيرسون - اللازمة لفحص ودراسة العلاقات بين المتغيرات بشكل منتظم. يمكن للعلماء بعد ذلك أن يحددوا مدى الترابط بين الخصائص القابلة للقياس. واسـتمر "تشـارلز سـبيرمان" Charles Spearman في تقليـد أسـلافه ومهد الطريق للتطوير اللاحـق وتعميم التحليل العامـلي لاحقـاً في أوائل القرن العشريـن. ومن الجدير بالذكر أن العديد من المسـاهمين الأوائل في القياس الرسـمي (بما في ذلك ألفريد بينيـه Alfred Binet، الذي طور اختبارات القدرة العقلية في فرنسا في أوائل القرن العشرين) قد تقاسموا الاهتمام بالقدرات الفكرية. وبالتالي، تم تطبيق الكثير من العمل المبكر في القياس النفسي على "الاختبار العقلي".

دور علم النفس الفسيولوجي:

نشـأ رافـدٌ تاريخي آخر للقياسـات النفسـية الحديثة من علم النفس الفسـيولوجي. وكما رأينا، كانت مشـاكل القياس شـائعة في علم الفلك والعلـوم الطبيعية الأخرى وكانت مصـدراً لاهتمام العالم إسـحاق نيوتن (بوتشـوالد، Buchwald, 2006). يقع علم النفس الفسـيولوجي عند التقاء علم النفـس والفيزياء ويهتم بالروابـط والصلات بين الخصائص

الفيزيائيــة للمحفزات أو المثيرات وكيف يدركها البشر وينظرون إليها. إن محاولات تطبيق إجراءات القياس المتبعة في الفيزياء على دراســة المشاعر (الأحاسيس) أدت إلى نقاش طويل حــول طبيعة القياس. وقد لخصت "نارينس ولــوسي" (Narens and Luce, 1986) هذه القضايا والمســائل. فقد لاحظ الباحثان أنه في أواخر القرن التاســع عشر، لاحظ هيلمهولتز (Helmholtz) أن الخصائص الفيزيائية (الطبيعية)، مثل الطول والكتلة، تمتلك نفس البنية الرياضية (mathematical) الحقيقية مثلها مثل الأعداد الحقيقية الموجبة[١]. على ســبيل المثــال، فإن وحدات الطول والكتلة يمكن ترتيبها والإضافــة إليها مثل الأعداد العادية. وفي أوائل القرن العشرين، اســتمر النقاش، واعتبرت الجمعية البريطانية لتقدم العلوم (للنهوض بالعلم) أن القياس الأساسي للمتغيرات النفسية يُعد أمراً مستحيلاً بسبب المشكلات الكامنة في ترتيــب أو إضافــة تصورات حســية أو إدراكية. جادل «ســتيفنز» (S. S. Stevens) أن قابليــة الجمع / الإضافــة، التي تنطبق على الطول أو الكتلة، ليســت ضرورية وأشــار إلى أنه يمكن للأفراد إصدار أحكام متســقة إلى حدٍ ما على شــدة (قوة) الصوت. على ســبيل المثــال، فالأفــراد يمكنهم الحكم على صوتٍ مــا بأنه أعلى من صوتٍ آخــر بمقدار الضعف أو النصف. وقال "ســتيفنز" إن خاصية النســبة هذه تُمكِّن من إخضاع البيانات المستمدة من هذه القياسات للمعالجة الرياضية. وينسب إلى "ستيفنز" تصنيف المقاييس إلى اسمية (nominal)، وترتيبيــة (ordinal)، وفترية (interval)، ونســبية (ratio)، وقال إن الحكم على قوة أو شــدة الصوت تتوافق مع مقياس النسبة (Duncan, 1984). وفي الوقت الذي كان فيه "ســتيفنز" يقدم حججه حول مشروعية قياس القياسات السيكولوجية، كان "ل.ل. ثورســتون" (L. L. Thurstone) يطور الأسس الرياضية للتحليل العاملي factor analysis (Nunnally, 1978). وقد امتدت اهتمامات "ثورســتون" لكل من القدرات العقلية وعلم

(١) مجموعة الأعداد الطبيعية تبدأ من الصفر إلى موجب ما لا نهاية بزيادة واحد صحيح في كل مرة. أما مجموعة الأعداد الصحيحة، فتشــتمل على الأعداد من سالب ما لا نهاية بالإضافة إلى الصفر بالإضافة إلى الأعداد الموجبة التي تحتويها مجموعة الأعداد الطبيعية بزيادة واحد صحيح كل مرة. أما الأعداد الكسرية، فتتكون من كسور الأعداد الصحيحة في صورة بسط ومقام. أما الأعداد الحقيقية، فتشمل المجموعات السابقة كلها بالإضافة إلى الأعداد التي لا يمكن كتابتها على شــكل كســور مثل باي (π) أي الأعداد اللاكسرية. ويمكن تصور الأعداد الحقيقية بأنها أعداد غير متناهية على خط مســتقيم. وتأخذ الأعداد الحقيقية اســمها من تضادها مع فكرة الأعداد التخيلية. كما يمكن لها أن تقوم بقياس الكميات المستمرة على اختلافها (المترجمَين).

النفــس الفســيولوجي. ووفقاً لدانكان (Duncan, 1984)، فإن "ســتيفنز" قد نســب إلى "ثورستون" تطبيق الطرق والأساليب السيكولوجية في قياس المحفزات الاجتماعية. وهكذا، فإن عمله يمثل تقارباً بين الجذور التاريخية المنفصلة.

التطورات اللاحقة في القياس:

تطور المفاهيم الأساسية:

وبقدر ما كان «ســتيفنز» مؤثرًا، فإن تصوره للقياس ليس هو الكلمة النهائية بأي حال مــن الأحوال. فقد عــرّف القياس بأنه «تخصيص وتعيين الأرقام للأشــياء أو الأحداث وفقًا للقواعد» (Duncan, 1984). وقد اعترض "دانكان" على هذا التعريف باعتباره غير مكتمل بنفس الطريقة التي يكون بها «تشــغيل البيانو عن طريق ضرب مفاتيح الآلة وفقًا لبعض الأنمــاط» غير مكتمل. فالقياس ليس فقط تعيين الأرقــام ...إلخ، بل هو أيضاً تعيين الأرقام بطريقة تتوافق مع درجات مختلفة من النوعية ... أو خاصية شيء أو حدث ما (ص ١٢٦).

وقــد حددت «نارينس ولــوسي» (Narens and Luce, 1986) أيضًــا قيودًا على تصور "ســتيفنز" الأصلي للقياس وأوضحا عددًا من التحسينات اللاحقة. ومع ذلك، يؤكد عملهم على نقطة أساســية أثارها "ســتيفنز" وهي: أنه توجد نماذج للقيــاس بخلاف النوع الذي أقرته اللجنة (الجمعية البريطانية لتقدم العلوم)، وتؤدي هذه النماذج إلى طرق قياس قابلة للتطبيق على العلوم غير الطبيعية وكذلك على العلوم الطبيعية. والخلاصة، إن هذا العمل المتعلق بالخصائص الأساســية للمقاييس قــد أرسى الشرعية العلمية لأنواع وأنماط إجراءات القياس المستخدمة في العلوم الاجتماعية.

تطور الاختبارات العقلية:

عــلى الرغم من أن الاختبارات العقلية (أو اختبــار القدرات، كما هو معروف الآن وأكثر شيوعًا)، كانت مجالًا تقليدياً نشطًا في القياس النفسي، إلا أنها لا تمثل محورًا أساسيًا لهذا الكتاب. ومع ذلك، فإنه تجدر الإشــارة إليها كمصدر للمساهمات الهامة في نظرية القياس وأساليبه. وقد نُشر لأول مرة في عام ١٩٦٨ منشور تاريخي حول النظريات الإحصائية لدرجات الاختبارات

العقليــة، قام بإعــداده «فريدريك م. لورد، وميلفين ر. نوفيــك» Frederic M. Lord and Melvin R. Novick وأعيــد إصداره مؤخرًا (Lord & Novick, 2008). وقد نشــأ وتطور هذا العمل من الأنشــطة الفكريــة الثرية لمجموعة الأبحاث الســيكومترية الخاصة بخدمة الاختبــارات التعليمية، والتــي قام بها "لورد ونوفيك". ويلخص هــذا النص المثير للإعجاب الكثــير مما كان معروفًا في مجال اختبار القدرات في ذلك الوقت وكان من بين أول الأوصاف المقنعة لما أصبح يعرف باســم نظرية اســتجابة البند (item response theory). وقد كان المدخل الأخير مناســبًا بشــكل خاص لمجال أوســع نطاقاً مثل الاختبارات العقلية. والعديد مــن أوجه التقــدم في هذا الفرع من القياس النفسي أقل شــيوعًا، وربما تكون أقل ســهولة في التطبيــق، عندمــا يكون الهدف هو قيــاس الخصائص الأخرى بخلاف القــدرات العقلية. ومع مرور الوقت، أصبحت إمكانية تطبيق هذه الطرق على ســياقات القياس بخلاف تقييم القــدرات أكــثر وضوحًا، وسنناقشــها في فصل لاحق. ومع ذلك، ســأركز في المقام الأول على الأساليب "الكلاسيكية أو التقليدية" التي سيطرت إلى حد كبير على قياس الظواهر الاجتماعية والنفســية غير القدرات. وهذه الطرق هي بصفة عامة أكثر مرونة وسلاســة بالنســبة لغير المتخصصين ويمكن أن تسفر عن نتائج ممتازة.

تقييم المرض العقلي:

كان لتطور أوصاف الأمراض العقلية تاريخ منفصل يقدم دراسة حاله مفيدة في كيفية أن عدم وجود نموذج قياس إرشــادي يمكن أن يجعل التقييم عمليه معقدة. وعلى مر القرون، تطــورت قــدرة المجتمع على التعرف عــلى أنواع مختلفة من الأمــراض العقلية عن طريق الملاحظــة غير المقننة تماماً نحو الجهــود المبذولة لفهم العلاقات بين الأعراض، والأســباب، والعلاجات المتوافقة مع القياسات الأكثر رسميه. لقد كانت رحله صعبة.

كانت الكتابات الرومانية واليونانية والمصرية المبكرة تساوي بين ما ندركه ونعرفه على أنه أعراض للمرض العقلي مع استحواذ الجن والشياطين أو غيرها من الظروف الخارقة للطبيعة «قوى فوق الطبيعة» (على ســبيل المثال، PBS, 2002). وبحلول عام ٤٠٠ قبل الميلاد، كان الطبيب اليوناني أبقراط (Hippocrates) يحاول فهم الظروف العقلية على أنها ناشــئة عن العمليات الفســيولوجية التي كانت محور التركيز الرئيسي في عمله العلمي (PBS, 2002).

وربمـا كانت جهوده من أوائل الجهـود التي بذلها في التفكير في المؤشرات الصريحة للمرض العقلي من حيث أسـبابه الكامنة. ومع ذلك، حتـى في تلك المرحلة وما بعدها، تم وصف الأمراض العقلية فينومينولوجياً[٢] (بشـكل ظاهري)؛ أي إن المظاهر المرتبطة بالمرض العقلي كانت مجرد بيانات وصفية بدلاً من فهمها كنقاط نهاية في تسلسـل له سبب أساسي واضح واحد أو أكثر من الأسباب الكامنة.

وقد اسـتمرت أسـاليب التصنيف البسيط للمرض العقلي لأكثر من ألف عام تقريباً. وقد لخـص «تارتاكوفسـكي» (Tartakovsky, 2011) كيـف تم تصنيف المـرض العقلي لأغراض التعداد في الولايات المتحدة في وقت مبكر من منتصف القرن التاسـع عشر. وفي تعداد عام ١٨٤٠، كانت هناك فئة واحدة فقط هي "البلاهة / الجنون" تشـير إلى وجود مشكلة عقلية. وبحلول عام ١٨٨٠، توسـع نظام تصنيف التعداد السـكاني ليشـمل الفئات التالية للمرض العقـلي: الهوس، والكآبة، والهوس الأحادي، والشـلل الجزئي، والخـرف، والهذيان، والصرع. وهذه هي الأوصاف الأساسية لحالات أو سلوكيات غير طبيعية (مثل الحزن المستمر أو الإفراط في شرب الخمر أو ضعف العضلات أو التشنجات) بدلاً من التصنيفات المسببة للأمراض.

وفي وقـت مبكر من ثمانينيات القرن التاسـع عشر، بدأ الطبيـب النفسي الألماني "إميل كريبيلين" (Emil Kraepelin) في التمييز بشكل أكثر منهجية بين الاضطرابات العقلية. وقد كان " كريبيلين" تلميذاً لعالم النفس الألماني "فيلهلم وندت" (Wilhelm Wundt) - والذي يُعتبر مؤسـس علم النفس التجريبـي - وكان "كريبيلين" طبيباً أيضـاً (Eysenck, 1968)، ومـن ثم قدم منظورين مختلفين لتصنيفاته للأمـراض العقلية. وفي عام ١٨٨٣، نشر كتاب باللغـة الألمانية تحت عنـوان (Compendium der Psychiatric) (Kraepelin, 1883) وهو كتاب أسـاسي يدعـو إلى تصنيف علمي بصـورة أكثر للأمراض النفسـية والتمييز بين

(٢) الفينومولجي (Phenomenology) الظاهراتية هي العلم الذي يدرس الظواهر، والمقصود من الظواهر هنا ليسـت ظواهـر العالم الخارجي، أي الظواهر الطبيعية الفيزيائية، بل المقصود بالظواهر ظواهر الوعي، أي ظهور موضوعات وأشـياء العالم الخارجـي في الوعي، وبذلك تكون الفينومينولوجيا هي دراسـة الوعي بالظواهـر وطريقة إدراكه لها وكيفيـة حضور الظواهـر في خبرته. فالفينومينولوجيا ليسـت مهتمة بما يصاحب عملية المعرفـة من حالات ذهنية أو وظائـف عضوية ترجع إلى الجهـاز العصبي، بل بكيفيـة إدراك الوعي للموضوع ووصولـه إلى معرفة موضوعية ويقينية حوله؛ أي بالاسـتعدادات المعرفية الموجودة لدى الذات الإنسـانية والتي تمكنها من تأسـيس معرفة يقينية، وهذه الاستعدادات ليست سيكولوجية، بل مرتبطة بالوعي الخالص قبل أن يتصل بأي خبرة تجريبية (المترجمَين).

الخرف بريكوكس (داء الجنون) والذهان الهوسى الاكتئابي. لكن مرة أخرى، على الرغم من جهوده لاستدعاء تفسـيرات لهذه الأمراض، فإن فئات التشخيص المبكرة الخاصة به هي في المقام الأول وصفًا موجزًا لأعراض واضحة تميل إلى الحدوث معاً، وليست التفسيرات المسببة المقنعة (Decker, 2007). على الرغم من أن "كريبلين" قد طور النهج العلمي لفهم المرض العقلي، إلا أن الأدوات المتاحة له كانت بدائية، وفي النهاية، كانت فئاته لتشـخيص الأمراض لا تزال وصفية إلى حد كبير.

ويُقيِّـم "ديكـر" (Decker, 2007) إرث "كريبيلين" على النحو التالي: "وخلاصه القول: وفقاً لمعايير البحث الحالية، فإن الأعمال العلمية لـ "كريبيلن" واسـتدلالاته من شـأنها أن تثير تساؤلات حول الأفكار والمفاهيم المسبقة وتحيز الملاحظين. ويمكن رؤية أوجه القصور العلمية في وصف "كريبيلين" نفسه لأساليبه ومنهجه. وعلى الرغم من تألقه في التصنيفات الفئوية الواضحة، فإن تراثه العلمي قام على أسس تجريبية هشة وضعيفة" (ص ٣٤١).

في منتصف القـرن العشرين، حاول الطب النفسي الأمريكي فرض المزيد من النظام على تقييـم المرض العقـلي. وبحلول موعد ظهور الدليل التشـخيصي والإحصـائي للاضطرابات العقليـة (DSM؛ الرابطة الأمريكية للطب النفـسي APA، ١٩٥٢)، حاولت نظم التصنيف السـائدة تصنيف الأمراض العقلية على أسـاس كل من مظاهرها ومسبباتها على حد سواء، كما هو الحال بالنسبة للصدمات الدماغية الحادة أو إدمان الكحول. ومع ذلك، لم يتم بعد تطوير مفاهيم أكثر دقة لمسببات الظروف غير المرتبطة بسبب خارجي واضح، وغالباً ما كان يفترض أسباب نفسية ديناميكية. وكان المصطلح المطبق على مثل هذه الحالات والظروف هـو ردود الفعل، التي يُفترض أنها ضغوط نفسـية ذات أصول غـير محددة. ومره أخرى، كانت التصنيفات في المقام الأول أوصافاً لأعراض ظاهرة. وعلى الرغم من أن نظام تصنيف الدليل التشخيصي والإحصائي للاضطرابات العقلية (DSM) يمثل تقدماً واضحاً يتجاوز النظم السابقة، إلا أنه لا يزال غير مطابق لمعايير القياس الحديثة. حتى بعد أربعه عقود منذ ذلك الوقـت، عندما ظهر الإصدار الرابع من الدليل التشـخيصي والإحصائي للاضطرابات العقلية DSM-IV (الجمعية الأمريكية للطب النفسي، ٢٠٠٠)، كان هناك قدر كبير من عدم الرضا عن نظام التصنيف. وقد لاحظ عالم النفس "بول ميهل" (Paul Meehl, 1999) أن المشكلة ليسـت بالضرورة في اسـتخدام الفئات (بعضها صعب وسريع، أو تنتمي أو لا تنتمي، ربما

كانــت هناك فئــات، كما قال) ولكن عــدم وجود مبرر منطقي واضــح لتعيين الناس لهذه الفئــات. وعــلى حد تعبــير "ميهل" (١٩٩٩)، "بالنســبة لتلــك الأقلية مــن قواعد الدليل التشخيصي والإحصائي للاضطرابات العقلية التي تشير إلى الكيانات التصنيفيةالحقيقية، فإن إجراءات تحديدها ومعايير تطبيقها تفتقر إلى أساس علمي كافٍ" (ص ١٦٦).

لقــد أتاحت مقدمة الإصدار الخامــس من الدليل التشــخيصي والإحصائي للاضطرابات العقليــة (DSM-V) وظهــوره في نهايــة المطــاف في ٢٠١٣ (الجمعيــة الأمريكيــة للطب النفــسي، ٢٠١٣) الفرصة لإعادة النظر في تصنيف الصحة العقلية. ويري البعض أن الفريق الذي عمل على التنقيح والمراجعة للدليل لم يســتفد اســتفادة كاملة من تلك الفرصة التي أُتيحت له. وكما لوحظ، فإن إحدى سمات تصنيف الصحة العقلية تاريخياً هي أنها سعت إلى التصنيف بدلاً من القياس. أي إن الهدف كان وصف الوجود أو الغياب بدلاً من تحديد درجة حالة محددة. وتشــير التجربة إلى أنه، حتى في الحــالات، مثل الفصام، فإن "ميهل" (١٩٩٩) كان على اســتعداد للاعــتراف بالتصنيف (أي أن تكون اضطرابات منفصلة ســواء كانت موجودة أو غائبة). فهناك سلسله متصلة من الضعف بدلاً من حالة الكل أو لا شيء. ومع ذلك لا يزال الاعتماد على التصنيف بدلاً من القياس قائماً.

في العديــد من الحــالات، كانت تنطوي على حدود تعســفية للعلامات والأعراض، مثل اجتياز خط وهمي من الشدة يشكل وجود حالة في حين أن عدم اجتياز هذا الخط الوهمي يعــبر عن عدم وجود الحالة. أيضًا، فقد اســتندت التصنيفات بشــكل حــصري تقريبًا إلى ملاحظات حول الأعراض الظاهرة بدلاً من تقييمات الدلالات الرئيسية للحالات، مثل وجود الأسباب المرضية السببية أو علامة وراثية أو حالة غير طبيعية من الكيمياء الداخلية التي قد تكون أساسًا لتعيين التشخيص الجسدي. عندما بدأ العمل (خارج الرؤية العامة) على الدليل التشــخيصي والإحصائي للاضطرابات العقلية الإصدار الخامسDSM-V، أعرب الكثيرون عن أملهــم في أن تكون المراجعــة والتنقيح أكثر جرأة من الإصدارات الســابقة وأن يتم تطبيق أســاليب تقييم أكثر حداثة. وفي عام ٢٠٠٥، بعد الإعلان عن خطط مراجعة وتنقيح الدليل التشــخيصي والإحصائي للاضطرابات العقلية (والتي ســتصبح الإصــدار الخامس من الدليل DSM-V)، بدأ المجتمع العلمي للصحة النفسية في التعبير عن مخاوفه. على سبيل المثال، ركز عدد خاص مــن مجلة علم النفس المــرضي (Journal of Abnormal Psychology)

عــلى أهمية وجدوى إعــادة التفكير في علم الأمراض النفســية اســتناداً إلى تحديد الأبعاد الأساســية، مثــل الفكر المضطــرب، والتأثير، والســلوك، التي تؤدي إلى مشــاكل محددة في الصحــة العقليــة (Kreuger, Watson, & Barlow, 2005). وقــد جادل كروجر وآخرون (Kreuger et al., 2005) بــأن هــذا المدخــل أو المنهــج يمكن أن يعالج اثنــين من أوجه القصور التجريبية الأساســية لنظم التصنيف القائمة على الفئة هما: انتشــار واسع للإصابة المرضية المشــتركة (أي مجموعات الأعراض الفردية التي تناســب التشــخيصات المتعددة) وعدم التجانس الشديد داخل التشخيص (أي الأفراد الذين تم تحديد نفس التشخيص لهم يتشــاركون في الأعراض بصورة قليلة أو ربما لا توجد أعراض). ولقد ضغط كل من الباحثين والمنظرين وحتى الفلاســفة (على سبيل المثال، أراغونا (٢٠٠٩) Aragona) من أجل إعادة النظر في تشــخيص المرض العقلي الذي كان أكثر انسجاماً مع العمل التجريبي مثل أساليب ومداخل القياس الحديثة. ومع ذلك، وعلى الرغم من هذه الجهود، فقد أصدرت الجمعية الأمريكية للأمراض النفسية الإصدار الخامس من الدليل التشخيصي والإحصائي للاضطرابات العقليــة DSM-V في شــكل احتفــظ بنظام التصنيف الأســاسي المســتخدم في الإصدارات الســابقة. وقد دفع ذلك "توماس إنســل" (Thomas Insel)، مدير المعهد الوطني للصحة العقليــة (NIMH)، إلى إصدار بيان على مدونته (Insel, 2013)، يقول: إن المعهد الوطني للصحــة العقليــة لم يعد ينظم جهــوده البحثية حول فئات الدليل التشــخيصي والإحصائي للاضطرابات العقلية وأنه بذل جهداً لمدة عشر ســنوات، وهو مشروع معايير مجال البحث (RDoC)، وذلك لإعادة صياغة مفهوم ووضع تصور للأمراض العقلية. وقد وصف إنســل (Insel, 2013) هــذا الجهد بالقول: إن مــشروع معايير مجال البحث "RDoC" هو إطار لجمع البيانات اللازمة لعلم تصنيف وتشــخيص الأمراض الجديد. ولكن من المهم أن ندرك أننا لا نستطيع النجاح إذا استخدمنا فئات الدليل التشخيصي والإحصائي للاضطرابات العقلية على أســاس أنه "المعيار الذهبي". وفي الشهر التالي، أصدر "إنسل" بيانًا صحفيًا مشتركًا مع الرئيــس المنتخب للجمعية الأمريكية للطب النفسي آنذاك "جيفري أ. ليبرمان" Jeffrey A. Lieberman. في هذا الإصدار، ذكروا ما يلي:

> اليوم، يمثل الدليل التشــخيصي والإحصــائي للاضطرابات العقلية (DSM) التابــع للجمعيــة الأمريكية للطب النفــسي (APA)، إلى جانــب التصنيف

الدولي للأمراض (ICD)، أفضل المعلومات المتاحة حاليًا للتشــخيص السريري للاضطرابات العقلية.

ومــع ذلك، فإن ما قد يكون ممكناً مــن الناحية الواقعية في الوقت الحاضر بالنســبة للممارســين لم يعد كافياً للباحثين. واستشرافاً للمســتقبل، فإن وضع الأساس لنظام تشــخيصي مســتقبلي يعكس بصوره مباشره علوم المخ الحديثة، سيتطلب الانفتاح لإعادة التفكير في الفئات التقليدية. ومن الواضح بشكل متزايد أن المرض العقلي سيتم فهمه على أفضل وجه على أنه اضطرابات في بنية الدماغ ووظيفتــه التي تنطــوي على مجالات محددة من الإدراك والعاطفة والســلوك. هذا هو التركيز الخاص أو محور مشروع معايير مجال البحث (RDoC) الخاص بالمعهد الوطني للصحة العقلية NIMH. (Insel & Lieberman, 2013).

في أكتوبر ٢٠١٥، اســتقال "إنسل" من منصبه في المعهد الوطني للصحة العقلية (إنسل، Insel, ٢٠١٥) لقبوله منصب في قسم علوم الحياة في شركة ألفا بيت (Alphabet)، والشركة هــي المظلة التي تشــكلت كجزء من إعادة التنظيم الهيــكلي لجوجل Google. وكان من بــين العوامل التي ذكرها "إنســل" أنها أثرت على قراره هــو أمله في التوصل إلى منهج أكثر تنظيــماً لتصنيف الصحة العقلية. كما ذكر في مقابلــة لمجلة إم.آي.تي لمراجعة التكنولوجيا (MIT Technology Review)، أن انتقاله إلى شركة "ألفا بيت"، في جزء منه، يمثل "محاولة لاكتشاف طريقه أفضل لجلب تحليلات البيانات إلى الطب النفسي. فنظام التشخيص الذي نملكه يستند كلياً على الأعراض القائمة وذات طابع شخصي إلى حدٍ ما" (Regalado, 2015). ويأمــل الكثــيرون في أن العمــل الذي يقوم به "إنســل" في شركة "ألفا بيت" ســوف يعزز تحديث التقييم النفسي لجعله أكثر توافقاً مع معايير القياس الحديثة.

توسيع نطاق القياس النفسي:

يذكــر دنــكان (Duncan, 1984) أن تأثير القيــاس النفسي في العلــوم الاجتماعية قد تجاوز أصوله في قياس الأحاسيس والقدرات الفكرية. ومن الواضح أن القياس النفسي ظهر كنمــوذج منهجي أو ميثودولوجي في حد ذاتــه. ويدعم "دنكان" هذه الحجة بثلاثة أمثلة لتأثير القياســات النفســية: (١) الاستخدام واسع الانتشــار للتعاريف السيكومترية للثبات

والصدق، (٢) انتشــار اســتخدام التحليل العاملي في أبحاث العلوم الاجتماعية، و(٣) تبنّي أساليب القياس النفسي من أجل تطوير المقاييس التي تقيس مجموعة من المتغيرات أوسع بكثير من تلك التي كانت لا تهتم بها القياســات النفســية في البداية (ص ٢٠٣). إن قابلية تطبيق المفاهيم والأســاليب الســيكومترية لقياس الظواهر النفســية والاجتماعية المتنوعة ستستحوذ على اهتمامنا فيما تبقى من هذا الكتاب.

دور القياس في العلوم الاجتماعية:

علاقة النظرية بالقياس:

الظواهــر التي نحاول قياســها في أبحاث العلوم الاجتماعية غالبــاً ما تنبع من النظرية. وبالتالي، تلعب النظرية دورًا رئيسيًا في كيفية تصورنا لمشاكل القياس الخاصة بنا. في الواقع، يذكر لــورد ونوفيك (Lord and Novick, 2008) أن للقضايا النظرية دوراً مهماً في تطوير نظريــة القياس. وكان المنظرون يشــعرون بالقلق من أن تقديــرات العلاقات بين المفاهيم البنائيــة constructs محــلاً للاهتمام، قد تــم الحصول عليها عمومًا مــن خلال المؤشرات المرتبطة بتلك المفاهيم البنائية. ونظراً لأن تلك المؤشرات تحتوي على خطأ، فإن الارتباطات الناتجــة عن ذلك كانت أقل من العلاقة الفعليــة بين المفاهيم البنائية، وقد حفز ذلك على تطوير واستحداث أساليب لتعديل وتصحيح الارتباطات من أجل تقليل الخطأ وحفز تطوير نظرية القياس كمجال اهتمام وتركيز متميز (ص ٦٩).

بطبيعة الحال، فإن العديد من مجالات العلوم تقيس الأشــياء المشــتقة أو المستمدة من النظرية. وحتى يتم تأكيد العناصر الفرعية من خلال القياس، فهي أيضًا مجرد مفاهيم بنائية نظريــة. ومع ذلك، تختلف النظريــة في علم النفس والعلوم الاجتماعية الأخرى عن النظرية في العلوم الطبيعية. ويميل العلماء في العلوم الاجتماعية إلى الاعتماد على العديد من النماذج النظريــة التــي تتعلــق أو تهتم بالظواهر الضيقــة النطاق والمحددة إلى حــدٍ ما، في حين أن النظريات في العلوم الطبيعية أقل من حيث العدد وأكثر شــمولًا من حيث النطاق أو المجال. على سبيل المثال، تركز نظرية المقارنة الاجتماعية لفستنجر (Festinger, 1954) على مجموعة ضيقة إلى حد ما من الخبرة الإنســانية: الطريقة التي يُقيِّم بها الناس قدراتهم أو آراءهم من

خلال مقارنة أنفسـهم مـع الآخرين. وعلى النقيض من ذلك، يواصـل علماء الفيزياء العمل نحـو إيجاد نظرية موحدة كبـيرة النطاق تضم جميع القوى الأساسـية للطبيعة ضمن إطار مفاهيمـي واحد. أيضاً، العلوم الاجتماعية أقل نضجاً مـن العلوم الطبيعية، ونظرياتها التي تتطور بسرعة أكبر. ويشـكل قياس الظواهر المعقدة وغير الملموسـة المستمدة من نظريات متعـددة ومتطورة، تحديًـا واضحًا للباحثين في العلوم الاجتماعيـة. ولذلك، من المهم بصفة خاصة أن نضع في اعتبارنا إجراءات القياس وأن ندرك تماماً نقاط قوتها وأوجه القصور فيها.

وكلما ازدادت معارف الباحثين عن الظواهر التي يهتمون بدراستها، والعلاقات التجريدية الموجودة بين المفاهيم البنائية الافتراضية، والأدوات الكمية المتاحة لهم، كانوا أكثر جاهزية بشكل أفضل لتطوير مقاييس تتميز بالثبات والصدق والقابلية للاستخدام والتطبيق. ولعل المعرفة التفصيلية بظاهرة محددة موضع الاهتمام تُعد أهم هذه الاعتبارات. على سـبيل المثال، نظرية المقارنة الاجتماعية لها العديد من الجوانب التي قد تنطوي على إستراتيجيات قيـاس مختلفة. وقد يتطلب أحد الأسـئلة البحثية تفعيل المقارنـات الاجتماعية باعتبارها تفضيلاً نسبياً للمعلومات المتعلقة بالآخرين ذوي المكانة الأعلى أو الأدنى، في حين قد يفرض سـؤال بحثي آخر تصنيف الذات بالنسبة إلى "الشخص العادي" على مختلف الأبعاد. وقد لا تسـفر المقاييس المختلفـة التي تقيس جوانب مختلفة من نفـس الظاهرة العامة (مثل المقارنة الاجتماعية) عن نتائج متقاربة (ديفيليس، DeVellis et al., 1990). وفي الحقيقة، تقوم المقاييس بتقييم متغيرات مختلفة على الرغم من استخدام اسم متغير مشترك أو عام في توصيفها. وبالتالي، فإن وضع وتطوير مقياس مناسـب على النحو الأمثل لسـؤال البحث يتطلب فهم التفاصيل الدقيقة للنظرية.

تتطلـب المتغـيرات المختلفة إسـتراتيجيات تقييـم مختلفة. على سـبيل المثال، يمكن ملاحظـة عدد العملات المميزة المأخوذة مباشرة من صندوق أو حاوية، في حين أنه لا يمكن ملاحظة الكثير من المتغيرات التي يهتم بها العلماء الاجتماعيون والسلوكيون بشكل مباشر؛ فالمعتقدات، والحالات التحفيزيـة، والتوقعات، والاحتياجات، والعواطف، وتصورات الدور الاجتماعي ليسـت سوى أمثلة قليلة. لا يمكن ملاحظة بعض المتغيرات مباشرة ولكن يمكن تحديدها بإجراءات بحثية غير الاسـتبيانات. على سـبيل المثال، على الرغم من أنه لا يمكن للباحثين المعرفيين أن يلاحظوا بشـكل مباشر كيفية تنظيـم الأفراد للمعلومات حول النوع

الاجتماعــي في مخططاتهــم الذاتيــة، إلا أنهم قد يكونون قادرين على اســتخدام إجراءات التذكر للاستدلال على كيفية تنظيم الأفراد لأفكارهم حول الذات والجنس. ومع ذلك، هناك العديد من الحالات التي يستحيل فيها أو لا يكون عملياً تقييم متغيرات العلوم الاجتماعية بأي طريقة أخرى غير اســتخدام مقياس القياس الورقي والقلم الرصاص. هذا هو الحال في كثير من الأحيان، ولكن ليس دائمًا، عندما يكون الاهتمام منصباً على قياس المفاهيم البنائية. وهكــذا، فــإن الباحث المهتم بقياس ظاهــرة أنثوية المظهــر androgyny قد يجد أنه من الأسهل بكثير القيام بذلك عن طريق استبانة تم تطويرها بعناية من خلال إجراءات بديلة.

المقاييس النظرية وغير النظرية:

في هذه المرحلة، ينبغي أن نعترف بأنه على الرغم من أن هذا الكتاب يركز على مقاييس المفاهيــم البنائية، فليس من الــضروري أن تكون كل التقييمات باســتخدام الورقة والقلم الرصــاص في حاجة إلى نظرية. فالجنس والســن، على ســبيل المثال، يمكــن التحقق منهما مــن التقرير الذاتي عن طريق اســتبانة. وبناءً على ســؤال البحث، يمكــن أن يكون هذان المتغــيران من مكونات نموذج نظري أو ببســاطة جزءاً من وصف المشــاركين في الدراســة. لا يوجد أساس نظري لبعض السياقات التي يُطلب فيها من الأشخاص الإجابة على قائمة من الأســئلة باستخدام نسق أو شــكل ورقي وقلم رصاص، مثل تقييم تفضيلات وجبة المريض في المستشــفى. وفي حــالات أخرى، قد تبدأ الدراســة غير نظرية ولكنهــا تؤدي إلى صياغة النظرية. على سبيل المثال، قد يطلب باحث في مجال التسوق من الآباء سرد أنواع الألعاب التي اشتروها لأطفالهم. بعد ذلك، قد يستكشف الباحث من هذه القوائم أنماط العلاقات. وبنــاءً على الأنمــاط الملاحظة لعمليات شراء الألعاب، يمكن للباحث تطوير نموذج لســلوك الشراء. وتمثل اســتبانات الرأي العام مثالاً آخر على القياس اللانظري إلى حدٍ ما. إن ســؤال الناس عن العلامة التجارية للصابون الذي يستخدمونه أو لمن يعتزمون التصويت لصالحهم في الانتخابــات نادراً ما ينطوي على أي محاولة للاســتفادة من مفهوم بنائي نظري أســاسي. وبــدلاً من ذلك، فإن الاهتمام يكمن في الاســتجابة الموضوعية في حد ذاتها، وليس في بعض خصائص الشخص التي يفترض أن يعبر عنها.

قد يكون التمييز بين مواقف القياس النظرية وغير النظرية أمرًا صعبًا في بعض الأحيان. على سبيل المثال، فإن البحث عن تفضيل الناخب في المرشحين للرئاسة كوسيلة للتنبؤ بنتيجة الانتخابات يعني مطالبة المجيب أو المجيبة بالإبلاغ عن نيته أو نيتها السـلوكية. قد يسأل الباحث الناس عن كيفية تخطيطهم للتصويت في الانتخابات، ليس بدافع الاهتمام بعمليات صنع القرار الانتخابي بل لمجرد توقع نتائج الانتخابات النهائية. ومن ناحية أخرى، إذا طرح السؤال نفسه في سياق دراسة حول كيفية تأثير الاتجاهات أو المواقف نحو قضايا محددة على تفضيل المرشحين، فإن وجود نظرية مفصلة بشكل جيد قد تكون أساساً للبحث. ولا تهدف المعلومات المتعلقة بالتصويت في هذه الحالة إلى الكشـف عن الكيفية التي سوف يصوت بها المستجيبون، ولكن لتسليط الضوء على الخصائص الفردية. في هاتين الحالتين، فإن صلة أو عدم صلة المقياس بالنظرية تُعد مسألة تتعلق بنية وهدف الباحث، وليست بالإجراءات المستخدمة. وتتم إحالة القراء المهتمين بمعرفة المزيد حول بناء استبانات المسح التي ليست معنيـة في المقام الأول بقياس المفاهيـم البنائية الافتراضية إلى المراجـع التالية: كونفيرس، وبيريـسر Converse and Presser (1986)؛ كزاجـا وبلـير Czaja and Blair (1996)؛ ديلـمان Dillman (2007)؛ فينـك Fink (1995)؛ فاولـر Fowler (2006)؛ وفيسـبرج، وكروسنيك، وبراون Weisberg, Krosnick, and Bowen (1996).

أنواع المقاييس:

غالبـاً ما يشـار إلى أدوات القيـاس التي تتكون مـن مجموعات من الفقـرات، والتي يتـم تجميعها معاً في درجه مركبة بهدف الكشـف عن مسـتويات المتغيرات النظرية التي لا يمكن ملاحظتها بسـهوله بالوسـائل المبـاشرة، على أنها مقاييس. ونحـن نطور مقاييس عندما نريـد قياس الظواهر التي نعتقد أنها موجودة بسـبب فهمنا النظري للعالم، ولكننا لا نسـتطيع أن نقيمهـا بصورة مباشرة. على سـبيل المثال، قد نقوم باسـتحضار الاكتئاب أو القلق كتفسـيرات للسـلوكيات التي نلاحظها. ويتفق معظم المنظرين على أن الاكتئاب أو القلـق لا يعادل السـلوك الـذي نراه ولكنه يكمـن وراءه. وتفـترض نظرياتنا أن هذه الظواهر موجودة وأنها تؤثر على السلوك ولكنها غير ملموسة. وفي بعض الأحيان، قد يكون من المناسـب اسـتنتاج وجودها من عواقبها السـلوكية. ومع ذلـك، في بعض الأحيان، قد

لا نتمكن من الوصول إلى المعلومات الســلوكية (كما هو الحال عندما نقتصر على استخدام منهجيات المســح البريــدي)، أو قد لا نكــون متأكدين من كيفية تفســير عينات أو نماذج الســلوك المتاحة (كما هو الحال عندما يظل الشــخص ســلبياً في مواجهة حدث قد يتفاعل معه معظم الأشخاص الآخرين بقوة)، أو قد لا يرغب في افتراض أن السلوك غير متماثل مع المفهوم البنائي الأســاسي محل الاهتمام (كما هو الحال عندما نشــك في أن البكاء هو نتيجة الفرح وليس الحزن). وفي الحالات التي لا نســتطيع فيها الاعتماد على السلوك كمؤشر على وجود ظاهرة ما، قد يكون من المفيد تقييم المفهوم البنائي من خلال مقياس تم بناؤه بعناية والتحقق من صدقه.

وحتــى بين المتغيرات المشــتقة من النظرية، هناك سلســلة متصلــة ضمنية تتراوح من ظواهر ملموسة نسبياً يمكن الوصول إليها إلى الظواهر المجردة نسبياً والتي يتعذر الوصول إليهــا. ولن يتطلــب كل شيء مقاييس متعددة البنود أو الفقــرات. ومن المؤكد أن العمر والجنس لهما صلة بالعديد من النظريات ولكن نادراً ما يتطلبان مقياسًا متعدد البنود لإجراء تقييم دقيق. يعرف الناس ســنهم وجنســهم. وترتبط هذه المتغيرات، في معظم الأحوال، بخصائص ملموسة لا لبس فيها نسبياً (مثل التشريح) أو الأحداث (مثل تاريخ الميلاد). وما لم تكن هناك ظروف خاصة مثل ضعف الجهاز العصبي، يمكن للمستجيبين استرجاع وتذكر المعلومــات المتعلقة بأعمارهم وجنســهم من الذاكرة بســهولة تامة. ويمكنهم الاســتجابة بدرجة عالية من الدقة على سؤال واحد لتقييم متغيرات مثل هذه. ويمكن القول إن الأصل العرقي أكثر تعقيدًا وتجريدًا من العمر أو الجنس، وعادةً ما ينطوي على مزيج من العوامل الماديــة والثقافية والتاريخية. ونتيجة لذلك، فإن المفهــوم البنائي الاجتماعي أقل في درجة الملموســية - من العمر أو الجنس. وعلى الرغــم من أن الآليات التي ينطوي عليها تحديد الشخص الأصل العرقي قد تكون معقدة وتتكشف على مدى فترة زمنية طويلة، فإن معظم الأفراد توصلوا إلى تعريف شخصي ويمكنهم الإبلاغ عن أصلهم العرقي بقليل من التفكير أو التأمــل. وبالتالي، قد يكفي متغير واحد لتقييم الأصل العرقي في معظم الظروف والحالات (غــير أن هذا قد يتغــير لأن مجتمعنا يصبح تدريجياً أكثر تعــدداً في العرقيات ولأن الأفراد يعرفــون انتماءهم العرقي الشــخصي من حيث المجموعات الاثنيــة المتعددة التي تعكس أســلافهم). ومع ذلــك، فإن العديد من المتغيرات النظرية الأخرى تتطلب من المســتجيب

إعادة بناء المعلومات القليلة المتاحة أو تفسيرها أو الحكم عليها أو مقارنتها أو تقييمها. على سبيل المثال، قياس كيف يعتقد الأشخاص المتزوجون أن حياتهم ستكون مختلفة إذا اختاروا زوجاً مختلفاً، على الأرجح يتطلب جهداً عقلياً وذهنياً كبيراً، وقد لا يستطيع بند واحد الكشف عن تعقيد الظاهرة محل الاهتمام. وفي ظروف كهذه، قد يكون المقياس هو أداه التقييم الأكثر ملاءمة. وقد تتمكن بنود متعددة من الكشف عن جوهر مثل هذا المتغير بدرجة من الدقة التي لا يمكن أن يحققها بند واحد. وهذا النوع من المتغيرات تحديداً - وهو المتغير الذي لا يمكن ملاحظته مباشرة والذي ينطوي على التفكير من جانب المستجيبين - هو الذي يتم تقييمه بشكل مناسب من خلال مقياس.

وينبغي مقارنة المقياس بأنواع أخرى من المقاييس المتعددة البنود التي تسفر عن درجة مركبة. والتمييز بين هذه الأنواع المختلفة من البنود المركبة ذات أهميه نظرية وعمليه على حد سواء، كما ستكشف الفصول اللاحقة. ووفقاً للمصطلحات المستخدمة في هذا الكتاب، فإن المقياس يتكون كما أشار «بولين» (65 – 1989, pp. 64؛ انظر أيضاً Loehlin, 1998, pp. 200-202) من "مؤشرات التأثير" effect indicators - وهي البنود التي تكون قيمتها ناتجة من المفهوم البنائي الأساسي (أو متغير كامن، كما سنشير إليها في الفصل التالي). وغالباً ما يتفق مقياس الاكتئاب مع خصائص المقياس، حيث إن الاستجابات على كل بند من البنود على حدة التي تتقاسم أو تشترك في قضية عامة - وتسمى الحالة الوجدانية للمستجيب. وهكذا، فإن كيفية استجابة شخص ما لبنود مثل "أشعر بالحزن" و"حياتي كئيبة" ربما تحددها إلى حد كبير مشاعر ذلك الشخص في ذلك الوقت. سأستخدم مصطلح «دليل» index، من ناحية أخرى، لوصف مجموعات من البنود التي تسبب مؤشرات - أي البنود التي تحدد مستوي مفهوم بنائي ما. فعلي سبيل المثال، من الممكن أن يكون قياس جاذبية المرشح الرئاسي مناسباً لخصائص الدليل index. وقد تقوم هذه البنود بتقييم مكان الإقامة الجغرافية للمرشح، وحجم أسرته، وجاذبيته المادية، وقدرته على إلهام العاملين في الحملة الانتخابية، والموارد المالية المحتملة. وعلى الرغم من أن هذه الخصائص ربما لا تشترك في أي قضية عامة، إلا أنها قد تتشارك جميعها في التأثير - مما يزيد من احتمال نجاح الحملة الرئاسية. فالبنود ليست نتيجة لشيء واحد، ولكنها تحدد النتيجة نفسها. وهناك مصطلح أكثر عمومية لمجموعه من البنود التي يمكن تجميعها في درجة مركبة هو المتغير الناشئ emergent variable (على سبيل المثال، Cohen, Cohen, Teresi, Marchi, & Velez, 1990)،

والذي يتضمن مجموعات من الكيانات التي تشــترك في خصائــص معينة ويمكن تجميعها تحت عنوان فئة مشــتركه. غير أن تجميعهم معاً لا يعني بالضرورة وجود أي علاقة ســببيه. فالجمل التي تبدأ بكلمة لها أقل من خمســة أحرف، على ســبيل المثال، يمكن تصنيفها بســهوله معاً على الرغم من أنها لا تشــترك في قضية مشــتركة ولا في أي أثر مشترك. إن المتغير الناشئ ينبثق لمجرد أن شخصاً ما أو شيئاً ما (مثل برنامج تحليل البيانات) يرى نوعاً من التشابه بين البنود المعنية.

جميع المقاييس ليست متساوية:

للأسف، لم يتم تطوير جميع البنود المركبة بعناية. وبالنسبة للكثيرين، قد يكون التجميع مصطلــح أكثر ملاءمة من التطوير. غالبًا مــا يقوم الباحثون بحذف البنود أو تجميعها معاً وافتراض أنها تشكل مقياسًا مناسبًا. قد لا يفكر هؤلاء الباحثون فيما إذا كانت البنود تشترك في قضية عامة (وبالتالي تشكل مقياسًا scale)، أو تشترك في إحداث نتيجة أو عاقبة مشتركة (وبالتالي تشكل دليل index)، أو مجرد أمثلة لفئة أهداف مرتبة مشتركة لا تنطوي على أي مقدمات سببية عامة أو نتيجة (وبالتالي تشكل متغيراً ناشئاً).

لا يفشــل الباحث في اســتغلال النظرية في تطوير المقياس فحسب، بل قد يتوصل أيضًا إلى اســتنتاجات خاطئة حول النظرية من خلال إســاءة تفســير ما يقيسه المقياس. وهناك أمر مؤســف ولكنه متكرر بشكل مقلق للغاية هو اســتنتاج أن بعض المفاهيم البنائية غير مهمــة أو أن نظرية معينة غير متناســقة بناءً على أداء مقياس قــد لا يعكس المتغير الذي يفترضــه الباحــث. لماذا قد يحدث هــذا؟ في البحوث نادراً ما نقوم بدراســة العلاقات بين المتغيرات بشكل مباشر. فكما ذكرنا سابقًا، فإن العديد من المتغيرات محل الاهتمام لا يمكن ملاحظتها مباشرة، وهذه حقيقة يمكننا نســيانها بســهولة. وفي كثير من الأحيان، فإننا نقوم بتقييم العلاقات بين ممثلين عن المتغيرات أو «وكلاء» proxies (مثل المقاييس) التي تهدف إلى تمثيــل متغــيرات محل الاهتمام. قد يصبــح هناك خلط بين الوكيــل أو الممثل proxy القابل للملاحظة والمتغير الذي لا يمكن ملاحظته. على سبيل المثال، يبدو أن متغيرات مثل ضغط الدم ودرجة حرارة الجســم، في الاعتبار الأول، يمكن ملاحظتهما بشكل مباشر، ولكن مــا نلاحظــه فعليًا عبارة عن وكلاء، مثل عمود الزئبق. تفترض اســتنتاجاتنا حول المتغيرات

أن الــوكلاء الذين تتم ملاحظتهم يرتبطون ارتباطًا وثيقًا بالمتغيرات الأساســية التي يمثلونها. هذا هو الحال بالنســبة لمقياس الحرارة؛ إننا نصف مســتوى الزئبق في مقياس الحرارة بأنه "درجــة الحرارة"، على الرغم من أنه بالمعنى الدقيق للكلمة، هو مجرد مظهر مرئي لدرجة الحــرارة (أي الطاقة الحراريــة). في هذه الحالة، حيث يتطابــق الطرفان عن كثب، تكون عواقــب الإشــارة إلى القياس (قيمة المقياس التي يصل إليها الزئبــق) كمتغير (كمية الطاقة الحرارية) تكاد تكون دائماً غير ذات أهمية. عندما تكون العلاقة بين المتغير ومؤشره أضعف مــما هي عليه في مثال مقياس الحرارة، فإن الخلط بــين المقياس والظاهرة التي يهدف إلى كشــفها يمكن أن يؤدي إلى اســتنتاجات خاطئــة. لنأخذ في الاعتبار حالــة افتراضية يرغب فيهــا الباحث في إجراء تحليل ثانوي لمجموعــة بيانات موجودة. ولنفترض أن هذا الباحث مهتــم بدراســة دور الدعم الاجتماعــي في التحصيل المهني اللاحــق. ويلاحظ الباحث أن مجموعة البيانات المتاحة تحتوي على ثروة من المعلومات حول الأوضاع المهنية للأشــخاص على مدى فترة زمنية طويلة وأن الأشــخاص قد تم ســؤالهم عما إذا كانوا متزوجين أم لا. في الواقــع، قد يكون هناك العديد من البنود التي تم جمعها في أوقات مختلفة، والتي تتعلق بالــزواج. ولنفترض كذلك، في حالة عدم وجود أي بيانــات توفر تقييماً أكثر تفصيلاً للدعم الاجتماعي، يقرر الباحث جمع بنود الزواج هذه في "مقياس" وأن يســتخدم ذلك كمقياس للدعم. ويتفق معظم علماء العلوم الاجتماعية على أن مســاواة الدعم الاجتماعي بالحالة الزواجيــة ليــس له ما يبرره. حيث يســقط هذا الأخير ويلغي جانبــين مهمين من جوانب الدعم الاجتماعي (على سبيل المثال، إدراك نوعية أو جودة الدعم الذي يتم الحصول عليه) ويتضمن عوامل يحتمل أنها غير ذات صلة بالموضوع (على ســبيل المثال، حالة طفل صغير جداً بحيث لا يتمكن من الزواج مقابل شخص بالغ في سن مناسب للزواج في وقت القياس). إذا اســتنتج هذا الباحث الافتراضي، اســتناداً إلى طريقة التقييم هذه، أن الدعم الاجتماعي لم يلعب أي دور في تحقيق التحصيل المهني، فإن هذا الاســتنتاج قد يكون خاطئًا تمامًا. في الواقــع، كانت المقارنة بين الحالة الزواجية والتحصيــل المهني (أو، على نحو أدق، مؤشرات هذين المتغيرين). ولا يكون الاستنتاج المتعلق بالدعم والتحصيل المهني صحيحاً وصادقاً إلا إذا أشار الزواج فعلاً إلى مستوي الدعم.

تكلفة سوء أو ضعف القياس:

وحتى إذا كان القياس السيئ هو المقياس الوحيد المتاح، فإن تكاليف استخدامه قد تكون أكبر من أي فوائد يمكن تحقيقها. والحالات نادرة في العلوم الاجتماعية التي يجب أن يُتخذ فيها قرار فوري من أجل تجنب العواقب الوخيمة، وليس أمام المرء خيار آخر سوى القيام باستخدام أفضل الأدوات المتاحة. ولكن حتى في هذه الحالات النادرة، لا تختفي وتتلاشى المشاكل المتأصلة في استخدام المقاييس الرديئة لتقييم المفاهيم البنائية. إن استخدام قياس لا يقوم بتقييم ما يفترضه المرء يمكن أن يؤدي إلى قرارات خاطئة. هل هذا يعني أننا ينبغي أن نستخدم فقط أدوات القياس التي خضعت للتطوير الدقيق واختبار الصدق الشامل؟ وعلى الرغم من أن القياس المعيب أو الناقص قد يكون أفضل من عدم وجود قياس على الإطلاق في بعض الحالات، فإنه ينبغي أن ندرك متى تكون إجراءات القياس الخاصة بنا بها عيوب وأن نقلل من استنتاجنا وفقاً لذلك.

في كثير من الأحيان، سوف يعتبر الباحث القياس كمسألة ثانوية لقضايا علمية أكثر أهمية والتي تكون حافزاً على الدراسة، وبالتالي، سيحاول الباحث أن يقتصد من خلال الحد من عملية القياس. ومع ذلك، فإن المقاييس المناسبة هي شرط ضروري لإجراء بحوث سليمة وصادقة. ويجب على الباحثين السعي لتحقيق التماثل بين المفاهيم البنائية النظرية محل اهتمامهم وطرق القياس التي يستخدمونها لقياسها إجرائياً. ويفرض سوء أو ضعف القياس قيوداً كبيرة على صدق الاستنتاجات التي يمكن الوصول إليها. بالنسبة للباحث الذي يفضل أن يولي قدراً ضئيلاً من الاهتمام للقياس وأكبر قدر ممكن للقضايا الموضوعية، فإن الإستراتيجية المناسبة قد تتمثل في تصحيح الجزء الخاص بالقياس منذ البداية المبكرة للبحث حتى يمكن اعتباره أمراً مفروغًا منه أو مسلماً به إلى حد ما بعد ذلك.

يمكن للباحث أيضًا أن يقتصد بشكل خاطئ من خلال استخدام المقاييس المختصرة جداً على أمل تخفيف العبء على المستجيبين. إن الاستبانة التي تكون مختصرة جدًا لكي تتمتع بالثبات تُعد فكرة سيئة بصرف النظر عن عدد المستجيبين الذين يفضلون الإيجاز. وينتج عن الاستبانة التي تتمتع بالثبات ويتم إكمال الإجابة عليها من قبل نصف المستجيبين معلومات أكثر من الاستبانة التي لا تتمتع بالثبات وتم الحصول على إجابة جميع المستجيبين عليها.

إذا لم تتمكــن من تحديد ما تعنيه البيانــات، فكمية المعلومات التي يتم جمعها تكون غير ذات صلة وغير مناســبة. وبالتالي، فإن اســتكمال الإجابة على الاســتبانات «الملائمة» التي لا يمكن أن تسفر عن معلومات ذات مغزى، هو استخدام سيئ لوقت المستجيبين وجهدهم مقارنةً باســتكمالهم نســخة أطول إلى حدٍ ما تنتج بيانات صادقة. وبالتالي، فإن اســتخدام أساليب وطرق تقييم مختصرة وغير كافية قد يكون لها آثار أخلاقية وكذلك آثار علمية.

الملخص والمراجعة:

يؤكد هذا الفصل على أن القياس نشاط أساسي في جميع فروع العلوم، بما في ذلك العلوم السلوكية والاجتماعية. وللقياس النفسي، وهو مجال التخصص في العلوم الاجتماعية المعني بقياس الظواهر الاجتماعية والنفســية، سوابق تاريخية تمتد إلى العصور القديمة. في العلوم الاجتماعيــة، تلعب النظرية دوراً حيوياً في تطوير القياس والمقاييس، والتي هي مجموعات من البنود التي تكشــف عن مســتوى متغير نظري أســاسي. ومع ذلك، لا تشــكل جميع مجموعات البنود مقاييس بهذا المعنى. وقد يكون تطوير المقاييس أكثر صعوبة من اختيار البنود بشــكل عرضي؛ ومع ذلك، فإن تكاليف استخدام المقاييس المبنية بشكل عرضي عادةً ما تفوق الفوائد والمنافع المتحققة إلى حد كبير.

تغطي الفصول التالية الأســاس المنطقي وأســاليب تطوير المقياس بمزيد من التفصيل. حيث يستكشــف الفصل الثاني المتغير الكامن، وهو المفهوم الأســاسي الذي يحاول المقياس قياسه كمياً، ويعرض الأسس النظرية للأساليب التي تم وصفها في الفصول اللاحقة. ويقدم الفصل الثالث أساساً مفاهيمياً لفهم الثبات والمنطق الكامن وراء معامل الثبات. ويستعرض الفصــل الرابع الصلاحية أو الصدق، في حين أن الفصــل الخامس هو دليل عملي للخطوات التي ينطوي عليها وضع وتطوير المقياس. ويقدم الفصل السادس مفاهيم التحليل العاملي ويصف اســتخدامها في تطوير المقياس. والفصل الســابع هو نظرة عامه مفاهيمية لمدخل بديل لتطوير المقياس - نظرية اســتجابة البند. وأخيراً، يناقش الفصل الثامن بإيجاز كيفية تناسب المقاييس مع عمليه البحث الأوسع نطاقاً.

تمارين:

١- ما هي الاختلافات الرئيسية بين المقياس (scale) والدليل (index) كما تم وصفهم؟

٢- هنـاك مهنتان لهـما تاريخ طويل في التقييم هما التعليم (من خلال تطوير واسـتخدام اختبـارات القـدرة الموحـدة) والطب النفسي (من خـلال مواصفـات وتطبيق معايير التشـخيص المعياري أو الموحد). ما هي بعض الاختلافات الرئيسـية بين كيفية اقتراب هذين المجالين البحثيين من مجال التقييم؟

الفصل الثاني
فهم المتغير الكامن

يقــدم هذا الفصل مخططًــا وتصوراً مفاهيميًــا لفهم العلاقة بــين المقاييس والمفاهيم البنائية constructs التي تمثلها، على الرغم من أنه ليس الإطار الوحيد المتاح. وتُعد نظرية استجابة البند منظور قياس بديل والذي سنقوم بدراسته في الفصل السابع. ونظرًا لأنه إطار مفاهيمي مهم ويمكن حسابه واستخدامه على نطاق واسع، فإنني أؤكد على نموذج القياس الكلاســيكي، الذي يفترض أن كل بند من البنود على حــدة هي مؤشرات قابلة للمقارنة مع المفاهيم البنائية الأساسية.

المفاهيم البنائية (constructs) مقابل المقاييس:

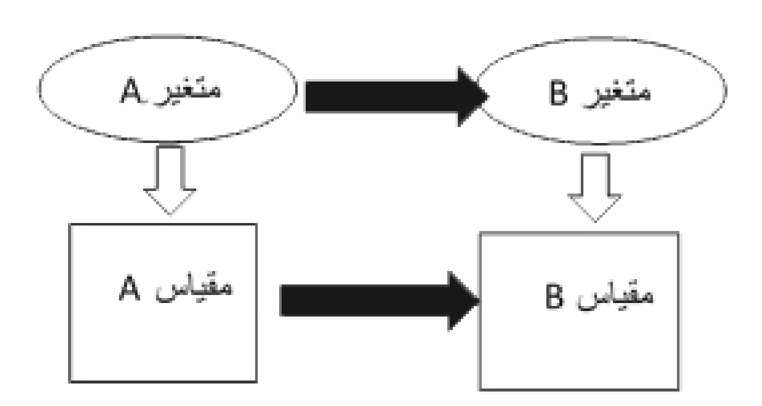
عــادةً مــا يهتم الباحثون بالمفاهيم البنائية بدلاً من البنود أو المقاييس في حد ذاتها. على ســبيل المثال، ســيكون باحث في مجال التســويق يهدف إلى قياس تطلعــات الوالدين تجاه أطفالهم أكثر اهتمامًا بالمشــاعر الأبوية غير الملموســة والآمال فيما سينجزه أطفالهم مقارنةً بالمكان الذي يضع فيه هؤلاء الآباء علامات على الاستبانة. ومع ذلك، فإن تسجيل الاستجابات على استبانة قد يكون، في كثير من الحالات، أفضل طريقة لتقييم تلك المشاعر والآمال. وعادةً ما تكون بنود المقياس وسيلة نهائية لتقييم المفهوم البنائي. وبعبارة أخرى، الاستبانة ضرورية لأن العديــد من المفاهيم البنائية لا يمكن تقييمها مبــاشرةً. بمعنى، أن المقاييس تعتبر وكلاء (proxies) للمتغيرات التي لا يمكننا ملاحظتها بشكل مباشر. ومن خلال تقييم العلاقات بين المقاييس، نســتنتج بشــكل غير مباشر العلاقات بين المفاهيم البنائية. في الشكل رقم (٢-١)، على ســبيل المثال، على الرغم من أن اهتمامنا الأســاسي هو العلاقة بين المتغيرين A وB، فإننا نقدر تلك العلاقة على أساس العلاقة بين المقاييس المقابلة لتلك المتغيرات.

وغالبًا ما تسمى الظاهرة الأساسية أو المفهوم البنائي الأساسي المراد أن يعكسه المقياس بالمتغير الكامن (latent variable). ما هو المتغير الكامن بالضبط؟ يكشف مسمى المتغير الكامن عن اثنين من الســمات الرئيســية. فبالنظر إلى المثال الخاص بتطلعات الآباء لما ســينجزه الأطفال.

أولاً، إن هذه التطلعات كامنة وليست ظاهرة أو واضحة. فطموحات الآباء لإنجازات أطفالهم لا يمكــن ملاحظتها بصــورة مباشرة. بالإضافة إلى ذلك، فإن المفهــوم البنائي الخاص بطموحات وتطلعــات الآباء لإنجازات أطفالهم هو متغير وليــس ثابت - أي إن بعض جوانبه متغيرة، مثل قوته (شــدته)، أو مقداره (أهميته). فقد تختلف تطلعات الآباء بشأن إنجازات أطفالهم وفقًا للوقت (على ســبيل المثال، خلال مرحلة طفولتهم مقابل مرحلة المراهقة)، والمكان (على سبيل المثال، في مجال رياضي مقابل الفصل الدراسي)، أو أشخاص (مثل الآباء الذين تختلف خلفياتهم أو وظائفهــم)، أو أي مزيج من هذه الأبعاد وغيرها. والمتغير الكامن هو الظاهرة الفعلية التي تكون موضع الاهتمام - في هذه الحالة، هو تطلعات الآباء لإنجازات الطفل.

وثمــة جانب آخر جديــر بالملاحظة في المتغــير الكامن، هو أنه عادةً ما يكون ســمة من سمات الفرد الذي هو مصدر البيانات. وهكذا، في مثالنا الحالي، فإن تطلعات الآباء هي سمة من ســمات الآباء وليس الأطفــال. ووفقًا لذلك، فإننا نقوم بتقييمــه من خلال جمع بيانات حــول معتقدات الآبــاء من الآباء أنفســهم. في حين أنه قد تكون هناك ظــروف يكون فيها شــكل ما من أشكال إعطاء المعلومات بالوكالة أو النيابة عن الغير (على سبيل المثال، مطالبة الوالدين بإعطاء معلومات عن بعض خصائص أطفالهم) أمراً مناســباً. بشــكل عام، سنطلب من المســتجيبين أن يقرروا ذاتياً المعلومات المتعلقة بأنفســهم. وعندما لا يكون الأمر كذلك، كما هو الحال في دراسة تشمل الآباء الذين يصفون تطلعات أطفالهم لأنفسهم، يجب توخي الحذر عند تفسير المعلومات الناتجة عن ذلك.

ويمكن القول، في هذه الحالة الافتراضية، بأن المتغير الكامن يمكن وصفه بشــكل أكثر دقة عــلى أنه تصورات الآباء لتطلعــات أطفالهم وليس تطلعات الأطفال في حد ذاتها. وبالمثل، إذا طلبنــا من مجموعة من المتســوقين تقييم خصائص متجر معين، فإننا نقــوم بتقييم تصورات المتسوقين بدلاً من جوانب المتجر نفسه (والتي يمكن تقييمها بسهولة أكبر عن طريق الملاحظة المباشرة). إن درجة أهمية التمييز بين تقييم تصورات المستجيبين فيما يتعلق ببعض المحفزات الخارجية (على سبيل المثال، التصورات المتعلقة بالمتجر)، مقارنةً بخصائص المحفزات الخارجية (على سبيل المثال، المتجر نفسه)، سوف تعتمد على الظروف والأهداف المحددة للتقييم. ومع ذلــك، وفي جميع الحالات، من المهم مراعاة هذا التمييز وتقديم تفســيرات مناســبة للبيانات الناتجة عن ذلك.



شكل (٢-١): تتطابق العلاقات بين الأدوات (المقاييس) مع العلاقات بين المتغيرات الكامنة فقط عندما يتطابق كل مقياس مع المتغير الكامن الخاص به

وعلى الرغم من أننا لا نستطيع أن نلاحظه أو نحدده كمياً بشكل مباشر، فمن المفترض أن المتغـير الكامن يأخذ قيمة محددة في إطار مجموعة محددة من الشروط. والمقصد من المقيـاس الذي تم تطويره لقيـاس متغير كامن هو تقدير مقـداره الفعلي في وقت ومكان القيـاس لكل شيء يتم قياسـه. هذا المقـدار الفعلي الذي لا يمكـن ملاحظته هو الدرجة الحقيقية.

المتغير الكامن كسبب مفترض لقيم البند:

ينطـوي مفهوم المتغير الكامن على وجود علاقة معيّنـة بينه وبين البنود التي تعبر عنه. ويُعتبر المتغير الكامن سبباً في درجة البند - أي قوة أو مقدار المتغير الكامن (أي قيمة درجاته الحقيقية) يُفترض أنها تتسبب في البند (أو مجموعة من البنود) للحصول على قيمة معينة.

ومـن الأمثلة التي قـد تعزز هذه النقطة: فيـما يلي مجموعة بنـود افتراضية لتقييم تطلعات الآباء لإنجازات الطفل:

١- إنجازات طفلي تحدد نجاحي.

٢- سأفعل أي شيء تقريبًا لضمان نجاح طفلي.

٣- ليست تضحية كبيرة إذا ساعدت طفلي على تحقيق النجاح.

٤- إنجازات طفلي أكثر أهمية بالنسبة لي من أي شيء آخر يمكنني التفكير فيه.

إذا تــم إعطــاء الفرصة للآباء للتعبير عن مدى موافقتهــم القوية على كل بند من هذه البنود، فإن تطلعاتهم الكامنة لإنجازات الطفولة ينبغي أن تؤثر على اســتجاباتهم. وبعبارة أخــرى، ينبغي لكل بنــد أن يبين مدى قوة المتغير الكامن (التطلعــات لإنجازات الأطفال). والدرجــات التي يتــم الحصول عليها في البند هي نتيجة ناجمة عــن قوة أو مقدار المتغير الكامن بالنسبة لذلك الشخص في ذلك الوقت بالذات.

وتنطوي العلاقة السببية بين متغير كامن ومقياس على علاقات تجريبية معينة. على سبيل المثــال، إذا كانــت قيمة البند ناتجة عــن متغير كامن، فينبغي أن يكــون هناك ارتباط بين هــذه القيمة والدرجة الحقيقية للمتغير الكامن. وكنتيجة لارتباط كل مؤشر من المؤشرات بالمتغــير الكامن، فإنــه ينبغي أن ترتبط هذه المؤشرات أيضًا مع بعضها البعض. ونظرًا لأننا لا نســتطيع تقييم الدرجة الحقيقية بشــكل مباشر، فإننا لا نستطيع حساب الارتباط بينها وبــين البند. ومع ذلك، عندما ندرس مجموعة من البنود التي يُفترض أنها ناتجة عن نفس المتغــير الكامن، فإنه يمكننا أن نفحص وندرس علاقاتها مع بعضها البعض. لذلك، إذا كانت لدينا عــدة بنود مثل تلك المذكورة أعلاه والتي تقيس طموحــات الآباء لإنجازات الأطفال، يمكننــا أن ننظر مبــاشرة في كيفية ارتباطها مــع بعضها البعض، ونســتدعي المتغير الكامن كأســاس للارتباطات بين البنود، وأن نســتخدم هذه المعلومات للاســتنتاج والاستدلال على مدى ارتباط كل بند ارتباطاً قوياً بالمتغير الكامن. وسأشرح بعد قليل كيف يمكن أن نتعلم ونستفيد من كل هذا من خلال الارتباطات المتبادلة بين البنود. أولاً، ومع ذلك، سوف أقوم بعرض بعض الإجراءات التخطيطية للمساعدة في جعل هذا التفسير أكثر وضوحاً.

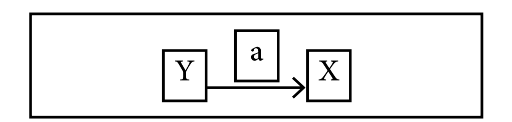
مخططات المسار:

ستقتصر تغطيه هذا الموضوع على استعراض موجز للقضايا ذات الصلة بتطوير المقياس. ولمزيد من التعمق في هذا الموضوع، يمكن الرجوع إلى «أشير» (Asher, 1983) أو "لوهلين" (Loehlin, 1998).

الاصطلاحات التخطيطية:

تعد مخططات المسار طريقة لوصف العلاقات السببية بين المتغيرات. وعلى الرغم من أنه يمكن استخدامها جنبًا إلى جنب مع تحليل المسار، وهو أسلوب تحليلي للبيانات، إلا أن مخططات المسار لها فائدة عامة أكبر كوسيلة لتحديد كيفية ارتباط مجموعة من المتغيرات ببعضها البعض. وتتقيد هذه المخططات ببعض الاصطلاحات. حيث يشير السهم المستقيم المرسوم من اسم متغير إلى اسم متغير آخر إلى أن المتغيرين الاثنين مرتبطان سببياً وأن اتجاه السببية هو ما يشير إليه السهم. وهكذا، فإن (X → Y) يشير صراحة إلى أن المتغير (X) هو سـبب في حدوث المتغير (Y). وفي كثير من الأحيان يتم تحديد المسارات الترابطية بواسطة علامات أو حروف، مثل الحرف (a) في الشكل رقم(٢-٢).

وغيابُ سـهمٍ ما لـه أيضًا معنى صريح، ألا وهو أن اثنين مـن المتغيرات لا توجد علاقة بينهما. وبالتالي، فإن (A → B → C D → E) تعني أن (A) يسبب (B)، و(B) يسبب (C)، وأن (C) و(D) لا توجد علاقة بينهما، وأن (D) يسبب (E).



شكل (٢-٢): المسار السببي من المتغير (X) إلى المتغير (Y)

وهناك اصطلاح آخر لمخططات المسـار وهو طريقة تمثيل الخطـأ (error)، والتي يتم تصويرهـا عادة كمتغير سـببي إضافي. ومصطلح الخطأ هذا عبارة عـن باقي (residual)، ويمثل كل مصادر التباين غير المحسوبة لأسباب أخرى مبينة بوضوح في الرسم التخطيطي.

ونظـرًا لأن مصطلح الخطأ هذا متبقـي، فإنه يمثل الاختلاف بين القيمة الفعلية للمتغير (Y) والقيمة التي نتوقعها للمتغير (Y) والمستندة إلى المعرفة بالمتغيرات (X) و(Z) (في هذه الحالة؛ انظر الشـكل ٢-٣). وفي بعض الأحيان، يتم افتراض مصطلح الخطأ، وبالتالي، لا يتم تضمينه في الرسم التخطيطي.

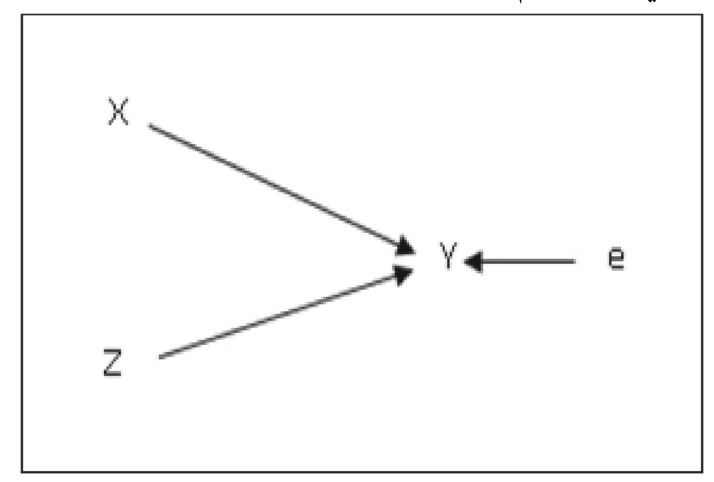
مخططات المسار في تطوير المقياس:

يمكن أن تساعدنا مخططات المسار في معرفه كيفية أو مدى ارتباط بنود المقياس بالمتغير الكامن سببياً. كما أنها يمكن أن تساعدنا أيضاً على فهم كيف أن بعض العلاقات بين البنود تنطوي على علاقات معينه بين البنود والمتغير الكامن. نبدأ بدراسة قاعدة حسابيه بسيطة لمخططات المسار. دعنا نلقي نظرة على الرسم التخطيطي للمسار البسيط في الشكل رقم (٢-٤).

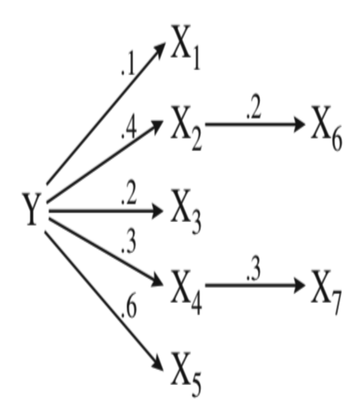
الأرقام على طول المسارات هي معاملات مسار معيارية (قياسية). وكل معامل مسار معياري يعبر عن قوه العلاقة السببية بين المتغيرات المرتبطة بالسهم. إن حقيقة كون المعاملات معيارية يعني أنها جميعاً تستخدم نفس المقياس (وحدات القياس) لقياس مقدار العلاقات السببية. في هذا الرسم التخطيطي، فإن المتغير(Y) هو سبب لكل من المتغيرات من (X1) إلى (X5). وتوجد علاقة هامة بين قيم معاملات المسار والارتباطات بين البنود Xs (التي تمثل البنود في حاله الرسم التخطيطي لمسار تطوير المقياس). بالنسبة للمخططات مثل هذه والتي يكون لها أصل مشترك واحد فقط (Yفي هذه الحالة)، فإن الارتباط بين أي اثنين من البنود (Xs) تساوي ناتج ضرب معاملات الأسهم التي تشكل مساراً، من خلال (Y)، بين المتغيرات(X) محل الاهتمام. على سبيل المثال، يتم حساب الارتباط بين (X1) و(X5) عن طريق ضرب معاملي المسارين المعياريين اللذَين يربطهما بـ (Y). وهكذا، فإن معامل الارتباط بينهما يساوي ($r_{1,5} = 0.6 \times 0.1 = 0.06$). كما تشترك المتغيرات (X6) و(X7) في (Y) كمصدر مشترك، لكن المسار الذي يربطهما يكون أطول. ومع ذلك، فإن قاعدة الحساب لا تزال سارية وتنطبق في هذه الحالة. فإذا بدأنا من (X7)، فإنه يمكننا تتبع العودة إلى (Y) ومن ثم التوجه إلى الأمام مرة أخرى إلى (X6) (أو في الاتجاه الآخر، من X6 إلى X7). وتكون النتيجة على النحو التالي: ($0.3 \times 0.3 \times 0.4 \times 0.2 = 0.0072$). وهكذا تكون قيمة ($r_{6,7}$) تساوي ٠,٠٠٧٢.

هذه العلاقة بين معاملات المسار والارتباطات توفر أساسًا لتقدير المسارات بين متغير كامن والبنود التي يؤثر عليها. وعلى الرغم من أن المتغير الكامن هو متغير افتراضي وغير قابل للقياس، إلا أن البنود حقيقية ويمكن حساب الارتباطات بينها بشكل مباشر. وباستخدام هذه الارتباطات، والقاعدة البسيطة للحساب التي تمت مناقشتها للتو، وبعض الافتراضات

حول العلاقات بين البنود والدرجات الحقيقية، يمكننا التوصل إلى تقديرات للمســـارات بين البنود والمتغير الكامن. ويمكننا أن نبدأ بمجموعة من الارتباطات بين المتغيرات. وبعد ذلك، بالعمل إلى الخلف أي العكس بأن نبدأ من العلاقة بين المسارات والارتباطات، يمكننا تحديد ما يجب أن تكون عليه قيم مســـارات معينة إذا كانت الافتراضات صحيحة. دعونا ننظر في المثال الوارد في الشكل رقم (٢-٥).



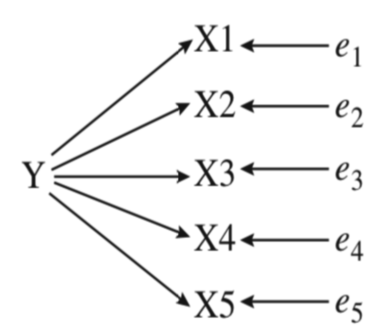
شكل (٢-٣): متغيران بالإضافة إلى تحديد الخطأ للمتغير (Y)



شكل (٢-٤): مخطط مسار ذو معاملات مسار والذي يمكن استخدامه لحساب الارتباطات بين المتغيرات

يشبه هذا الرسم التخطيطي المثال الذي تم تناوله سابقاً في الشكل رقم (٢-٤) باستثناء أنه لا توجد قيم مسـارات، وتم حذف وإسـقاط المتغيرين (X6) و(X7)، وتمثل المتغيرات (X) المتبقيـة بنود المقيـاس، ولكل بند متغير (خطأ error) يؤثـر فيه بخلاف المتغير (Y). ومتغـيرات الخطأ (e)هذه هـي متغيرات فريدة من نوعها في حالـة كل بند وتمثل التباين المتبقي في كل بند لم يتم تفسيره بواسطة المتغير(Y). ويشير هذا المخطط إلى أن كافة البنود تتأثـر بالمتغير(Y). بالإضافة إلى ذلك، يتأثر كل منها بمجموعه فريدة من المتغيرات بخلاف المتغير(Y) والتي يتم التعامل معها بشكل جماعي على أنها خطأ (error).

ويمثل هذا المخطط المعد لكيفية ارتباط خمسـة بنود كل على حدة بمتغير كامن واحد (Y). وتشـير الأرقـام العددية المعطاة لمتغـيرات الخطأ (es) وبنود المقيـاس (Xs) إلى أن البنود الخمسـة مختلفة وأن المصادر الخمسـة للخطأ، واحد لكل بند، هي أيضاً مختلفة. ولا يحتوي المخطط على أسهم للانتقال مباشرةً من بند (X) إلى بند آخر (X) أو الانتقال من متغير خطأ (e) إلى متغير خطأ آخر (e) أو من متغير خطأ (e) إلى بند (X) غير تلك المرتبطة بها. وتمثل هذه الجوانب من المخطط الافتراضات التي ستتم مناقشتها لاحقاً.



شكل (٢-٥): مخطط مسار يتضمن متغيرات الخطأ

إذا كانـت لدينا خمسـة بنـود فعلية أكملت الإجابـة عليها مجموعة من الأشـخاص، فسـتكون لدينا درجات بند يمكن أن نربطها مع بعضها بعد ذلك. ولقد سـمحت القاعدة التي تمت دراسـتها في وقت سـابق بحسـاب الارتباطات من معاملات المسار. ومع إضافة بعض الافتراضات، فإنه يتيح لنا أيضًا حسـاب معاملات المسار من الارتباطات، أي إنه يمكن

استخدام الارتباطات المحسوبة من البنود الفعلية لتحديد مدى ارتباط كل بند بالمتغير الكامن. على سبيل المثال، إذا كان الارتباط بين (X1) و(X4) يساوي (٠,٤٩)، فعندئذ نعرف أن ناتج قيم المسار الذي يمتد من (Y) إلى (X1) والمسار الذي يمتد من (Y) إلى (X4) يساوي (٠,٤٩).ونحن نعرف هذا لأن قاعدتنا الحسابية أثبتت أن الارتباط بين متغيرين يساوي ناتج ضرب معاملات المسار على طول المسار الذي يربطهما معاً. وإذا افترضنا أيضًا أن قيمتَي المسار متساوية، فيجب أن يكون كلاهما (٠,٧٠)[١].

مزيد من التوسع في نموذج القياس:

افتراضات القياس الكلاسيكية (التقليدية):

إن نموذج القياس الكلاسيكي أو التقليدي - الذي يؤكد أن الدرجة الملاحظة (X)، الناتجة عن حاصل جمع الدرجات الحقيقية[٢] (T)، بالإضافة إلى خطأ القياس (e) - تبدأ بافتراضات عامة حول البنود وعلاقاتها بالمتغير الكامن ومصادر الخطأ، وهذه الافتراضات هي:

١- يختلف مقدار الخطأ المرتبط بكل بند من البنود على حدة بشكل عشوائي. ويكون للخطأ المرتبط بكل بند من البنود على حدة متوسط يساوي (صفراً) عند تجميعه عبر عدد كبير من الأشخاص. وبالتالي، فإن متوسطات البنود تميل إلى عدم التأثر بالخطأ عندما يكمل عدد كبير من المجيبين الإجابة على البنود.

٢- لا يرتبط متغير الخطأ (error term) لأي بند مع متغير الخطأ لبند آخر. أما المسارات الوحيدة التي تربط البنود فهي تمر دائماً عبر المتغير الكامن، ولا تمر أبداً عبر أي متغير للخطأ.

٣- لا ترتبط متغيرات الخطأ مع الدرجة الحقيقية للمتغير الكامن. لاحظ أن المسارات المنبثقة من المتغير الكامن لا تمتد للخارج إلى متغير الخطأ. ويستهدف السهم بين بند ما ومتغير الخطأ الخاص الاتجاه الآخر.

---

(٢) الدرجة الحقيقية هي مفهوم افتراضي، يشير إلى خصائص لا يمكن ملاحظتها، ولا يمكن قياسها بشكل مباشر. وتُعرف إجرائياً بأنها متوسط درجات الفرد المتوقعة على سمة أو متغير معين عند قياس السمة أو المتغير عدداً لا نهائياً من المرات. وبسبب عدم إمكانية إعادة تطبيق الاختبار على الفرد إلى ما لا نهاية فإن إمكانية التحديد التام للدرجة الحقيقية يعد أمراً تقريبياً (المترجمَين).

الافتراضـــان الأول والثـــاني أعلاه هما افتراضَين إحصائيَّين شـــائعين تكمنان أو تقفان وراء العديـــد من الإجراءات التحليلية. أما الافتراض الثالث فهو تعريف "الخطأ" على أنه الباقي أو المتبقي بعد الأخذ في الاعتبار كل العلاقات بين مجموعه من المنبآت والنتيجة أو، في هذه الحالة، مجموعه من البنود ومتغيرها الكامن.

الاختبارات المتكافئة (المتوازية):

تعتمـــد نظرية القياس الكلاســـيكية، في شـــكلها الأكثر تقليدية، على افـــتراض اختبارات التكافؤ أو التوازي parallel tests. وينبع مصطلح "اختبارات التكافؤ" من حقيقة أنه يمكن للمـــرء أن ينظـــر إلى كل بند على حدة على أنه "اختبـــار" لقيمة المتغير الكامن. ولأغراضنا، فإن الإشـــارة إلى البنود المتكافئة ســـتكون أكثر دقة. ومع ذلك، سوف أؤجل استخدام ذلك وأستخدم الاسم التقليدي.

ومـــن مزايا نموذج اختبارات التكافؤ أن افتراضاته تجعل من الســـهل للغاية التوصل إلى اســـتنتاجات مفيدة حول كيفية ارتباط كل بندين من البنود على حدة بالمتغير الكامن بناءً على ملاحظاتنا عن كيفية ارتباط البنود ببعضها البعض. في وقت سابق، اقترحت أنه بمعرفة الارتباطات بين البنود وفي ظل افتراضات معينة، يمكن للمرء عمل اســـتدلالات واستنتاجات بشـــأن المســـارات المؤدية من متغير ســـببي إلى بند ما. كما هو مبين في الفصل التالي، فإن القـــدرة على تعيين وتحديد قيمة رقمية للعلاقات بين المتغير الكامن والبنود نفســـها مهمة جدًا. ومن ثم فإنني، في هذا القســـم، ســـأبحث بشيء من التفصيل كيف تؤدي افتراضات اختبارات التكافؤ إلى استنتاجات معينة تجعل ذلك ممكنًا.

والأســـاس المنطقي الذي يقوم عليه نموذج الاختبارات المتكافئة هو أن كل بند من بنود المقيـــاس هو عـــلى وجه التحديد مقياس جيـــد للمتغير الكامن مثل أي بنـــد آخر من بنود المقياس. وبالتالي، فإن كل بند من البنود تكون متكافئة تماماً، أي إن علاقة كل بند بالمتغير الكامـــن يُفترض أنها متطابقة مـــع علاقة كل بند آخر بهذا المتغير ويُفـــترض أيضاً أن يكون مقدار الخطأ الموجود في كل بند متطابقًا أيضًا. وبشكل تخطيطي، يمكن تمثيل هذا النموذج على النحو المبين في الشكل رقم (٢-٦).

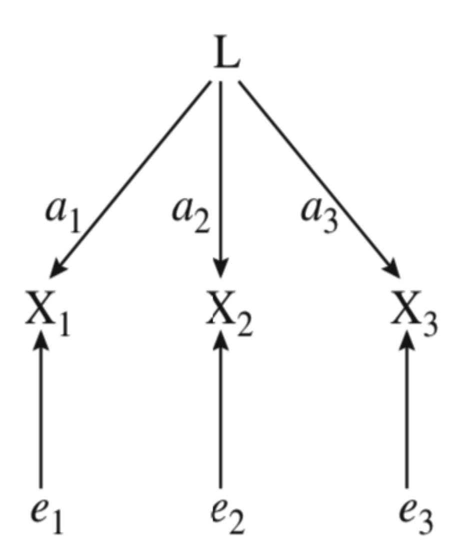
ويضيف هذا النموذج افتراضين إلى تلك الافتراضات المذكورة آنفاً وهما:

١- يفترض أن يكون مقدار التأثير من المتغير الكامن لكل بند هو نفسه بالنسبة لجميع البنود.

٢- يفــترض أن يحتــوي كل بند نفس مقدار الخطأ لأي بند آخر، مما يعني أن تأثير العوامل الأخرى غير المتغير الكامن متساوٍ لكافة البنود.

تعنــي هذه الافتراضات المضافة أن ارتباطات كل بند والدرجة الحقيقية تكون متطابقة. والقدرة على التأكيد بأن هذه العلاقات متساوية هو أمر مهم لأنها تؤدي إلى وجود وسيلة لتحديد قيمه كل من هذه الارتباطات المتطابقة أو المتماثلة، وهذا بدوره يؤدي إلى طريقة لقياس مدى الثبات، والذي ستتم مناقشتها في الفصل التالي.

إن التأكيــد عــلى ان الارتباطات بين الدرجــة الحقيقية وكل بند متســاوية يتطلب كل الافتراضات الســابقة. ومربع الارتباط هو نســبه التباين المشــترك بين متغيرين. لذلك، إذا كانت الارتباطات بين الدرجة الحقيقية وكل بند من بندين متســاوية، فإن نســب التباين المشــترك بين الدرجة الحقيقية وكل بند يجب أن تكون متســاوية أيضاً. افترض أن الدرجة الحقيقيــة تســهم بنفس مقدار التباين لــكل بند من البندين. فإن هــذا المقدار يمكن أن يكون نســبه متســاوية من التباين الكلي لكل بند فقط إذا كانت البنــود لها تباينات كلية متطابقة. ولكي تكون التباينات الكلية متساوية للبندين، فإنه يجب أن يكون مقدار التباين الــذي يتلقــاه كل بند من مصادر أخرى غير الدرجة الحقيقية متســاوٍ أيضاً. ونظرًا لأن كل مصــادر التباين بخــلاف الدرجة الحقيقية يتم تجميعها معًا أنهــا خطأ، فإن هذا يعني أنه يجب أن يكون للبندين تباينات خطأ متســاوية. على ســبيل المثال، إذا حصلت (X1) على تســع وحدات عشــوائية من التباين من درجتها الحقيقية ووحــدة واحدة من الخطأ، فإن نسبة الدرجة الحقيقية ستكون (٩٠٪) من التباين الكلي. إذا حصلت (X2) أيضًا على تسع وحدات تباين من الدرجة الحقيقية، فإن هذه الوحدات التســع يمكن أن تكون (٩٠٪) من الإجمالي فقط إذا كان التباين الكلي يســاوي (١٠). ولا يمكن أن يســاوي المجموع القيمة (١٠) إلا إذا ســاهم الخطأ بوحدة واحدة في (X2) مثلما ســاهمت في (X1). عندئذٍ، فإن الارتباط بين كل بند والدرجة الحقيقية يســاوي الجذر التربيعي لنســبة تباين كل بند الذي يعزى إلى الدرجة الحقيقية، أو ما يقرب من (٠,٩٥) في هذه الحالة.



شكل (٢-٦): مخطط لنموذج الاختبارات المتكافئة، حيث تكون جميع المسارات من المتغير الكامن (L) إلى البنود (X1, X2, X3) متساوية في القيمة مع بعضها البعض، وكذلك جميع المسارات من متغيرات الخطأ إلى البنود

وبالتالي، نظرًا لأن نموذج الاختبارات المتكافئة يفترض أن مقدار التأثير من المتغير الكامن هو نفسه لكل بند وأن المقدار من المصادر الأخرى (الخطأ) هي نفسها لكل بند، فإن نسب تباين البند المنسوبة إلى متغير كامن والخطأ يكونان متساويان بالنسبة لجميع البنود. وهذا يعني أيضًا أنه وفقًا لافتراضات الاختبارات المتكافئة، تكون معاملات المسار المعيارية من المتغير الكامن لكل بند متساوية لكافة البنود. وكان من المفترض أن تكون معاملات المسار المعيارية متساوية مما أتاح لنا إمكانية، في مثال سابق، حساب معاملات المسار من الارتباطات بين البنود. يجب أن تساعدنا قاعدة مخطط المسار المتعلقة بمعاملات المسار مع الارتباطات، والتي تمت مناقشتها سابقًا، في فهم سبب استمرار ووجود هذه المعادلة أو التكافؤ عندما يقبل المرء الافتراضات السابقة.

كما أن افتراضات هذا النموذج تعني ضمنياً أن الارتباطات بين البنود تكون متطابقة أو متماثلة (على سبيل المثال، الارتباط بين (X1) و(X2) مطابق للارتباط بين (X1) و(X3) أو بين (X2) و(X3). كيف نتوصل إلى هذا الاستنتاج من الافتراضات؟ الارتباطات كلها

متماثلـة لأن الآلية الوحيدة لحسـاب الارتباط بين أي بندين هي المسـار من خلال المتغير الكامـن الذي يربط بين هذين البندين. على سـبيل المثال، يتم ربـط (X1) و(X2) فقط بواسـطة المسار المكون من المسارين (a1) و(a2). ويمكن حساب الارتباط عن طريق تتبع المسـار الذي يربط بين البندين المعنيين وحاصل ضرب قيم المسار. بالنسبة لأي بندين من البنـود، يسـتلزم ذلك ضرب مسـارين لهما قيم متطابقـة (أي، a1 = a2 = a3). وبطبيعة الحال، ستكون الارتباطات المحسوبة من خلال ضرب القيم المتساوية متساوية.

كما ان الافتراضات تعني أن كل من هذه الارتباطات بين البنود تساوي مربعاً أي مساراً مـن المتغـير الكامن إلى بند من البنـود على حدة. كيف نصل إلى هذا الاسـتنتاج؟ حاصل ناتج ضرب مسـارين مختلفين (على سـبيل المثـال، a1 وa2) يكون مطابقـاً لمربع أي من المسارين لأن معاملات كلا المسارين متماثلان. إذا كانت a1 = a2 = a3، وكان حاصل ضرب (a1 x a2) = حاصل ضرب (a1 x a3) = حاصل ضرب (a2 x a3)، فعندئذٍ يجب أن يساوي كل مـن هذه النواتج الأخيرة أيضًا قيمة أي من المسـارات (a) مضروباً في نفسـه أي مربع المسـار (a). وقد يؤدي النظر إلى الوراء في الشـكل رقـم (٢-٦) إلى توضيح هذه العلاقات والآثار المترتبة عليها بشكل أكبر.

ويسـتنتج أيضاً من افتراضات هذا النموذج أن نسبة الخطأ المرتبطة بكل بند هو مكمل لنسـبة التباين المرتبطة بالمتغير الكامن. وبعبارة أخـرى، فإن أي تأثير على بند معين لا يتم تفسيره بواسطة المتغير الكامن يجب أن يتم تفسيره بواسطة متغير الخطأ (error). ويفسر ويـشرح هذين التأثيرين معاً (١٠٠٪) من التباين في أي بند معين. وهذا ببسـاطه لأنه يتم تعريف متغير الخطأ (e) على أنه يشمل جميع مصادر التباين في البند بخلاف المتغير الكامن.

وتدعم هذه الافتراضات اسـتنتاجاً واحداً آخر على الأقل؛ نظراً لأن كل بند يتأثر بنفس المقدار بالمتغير الكامن كما أن تأثير كل متغير من متغيرات الخطأ على البند المقابل له متساوٍ أيضاً، فإن كافة البنود تكون لها متوسطات متساوية وتباينات متساوية. إذا كان المصدران الوحيـدان اللذان يمكـن أن يؤثرا على المتوسـطات هما مصدران متماثـلان أو متطابقان بالنسـبة لكافة البنود، فمن الواضح أن المتوسـطات الخاصة بكافة البنود ستكون متطابقة أيضاً. وينطبق هذا المنطق أيضاً على التباينات الخاصة بالبنود.

وفي الختام، فإن نموذج الاختبارات المتكافئة أو المتوازية يفترض ما يلي:

١- عشوائية الخطأ.

٢- عدم ارتباط الأخطاء مع بعضها البعض.

٣- الأخطاء غير مرتبطة بالدرجة الحقيقية.

٤- يؤثر المتغير الكامن على جميع البنود بنفس القدر.

٥- مقدار الخطأ لكل بند يكون متساوياً.

تسمح لنا هذه الافتراضات أن نتوصل إلى مجموعه متنوعة من الاستنتاجات المثيرة للاهتمام. وعلاوة على ذلك، يتيح لنا النموذج استخلاص استنتاجات حول المتغير الكامن بناءً على ارتباطات البنود مع بعضها البعض. ومع ذلك، فإن النموذج يحقق هذا العمل من خلال وضع افتراضات دقيقة وصارمة إلى حدٍ ما.

نماذج بديلة:

وكما يحدث، فإن جميع الافتراضات المقيدة بقوة والمرتبطة بالاختبارات المتكافئة أو المتوازية التامة ليست ضرورية من أجل تحقيق استنتاجات مفيدة حول علاقة الدرجات الحقيقية بالدرجات الملاحظة. إن النموذج القائم على ما يسمى تقنياً باختبارات «تاو» المتكافئة (tau-equivalent) يجعل أحد الافتراضات أكثر تحرراً - وهو أن مقدار التباين في الخطأ المرتبط ببند معين لا يلزم أن يساوي تباين الخطأ في البنود الأخرى (على سبيل المثال، "الين، وين" Allen& Yen, 1979). ولكن لا تزال اختبارات "تاو" المتكافئة تتطلب درجات حقيقية متطابقة للبنود، على الرغم من أن التخفيف البسيط من هذا الافتراض يُعرف باسم اختبارات "تاو" المترادفة أو المتكافئة في الأساس **essentially tau-equivalent tests** (أو في بعض الأحيان اختبارات متكافئة عشوائياً **randomly parallel tests**). قد يكون لأي زوج من البنود الخاضعة أو المقيدة باختبارات "تاو" المترادفة أو المتكافئة في الأساس درجات حقيقية تختلف باختلاف الثابت. وبطبيعة الحال، فإن إضافة ثابت إلى بند واحد لا يكون له تأثير على أي ارتباط ينطوي على ذلك البند، لأن الارتباطات عبارة عن مقادير معيارية. وبالتالي، فإن العلاقة بين أي زوج من البنود أو بين الدرجة الحقيقية للبند والدرجة التي

حصــل عليها البند لا تتأثــر بالتخفيف من الافتراضات الخاصة باختبــارات «تاو» المتكافئة الصارمة مع تلك الافتراضات الخاصة بنموذج «تاو» المترادف أو المتكافئ في الأســاس. إذن، ما تم قوله حتى الآن حول اختبارات «تاو» المتكافئة ينطبق أيضًا على نموذج «تاو» المترادف أو المتكافئ في الأســاس. في أي من هاتين الحالتين، قد لا تكون القيم المعيارية للمســارات من المتغير الكامن لكل بند متساوية. ومع ذلك، فإن القيم غير المعيارية للمسار من المتغير الكامن إلى كل بند (أي، المقدار المقابل لنســبة التأثــير التي يمتلكها المتغير الكامن على كل بنــد) يفترض أنها لا تزال متطابقة بالنســبة لكافة البنود. وهذا يعنــي أن البنود متكافئة أو متوازيــة فيما يتعلق بمقدار تأثرها بالمتغير الكامن ولكنها لا تتأثر بالضرورة بنفس القدر مــن العوامــل الخارجية التي يتم تجميعهــا معًا كخطأ. وفي ظل افتراضــات التكافؤ التام، لا تقــوم بنود مختلفة فقط بالتأثير على الدرجة الحقيقية بنفس القدر أو الدرجة؛ بل أيضاً يكــون لمكونات الخطأ الخاصة بهم نفس التأثير. وتُعد مكافئ «تاو» (تاو"tau"هي المقابل اليونــاني لـ τ، كما هو الحال في الدرجة الحقيقية) أســهل بكثــير في التعامل لأنه لا يفترض شرط "تساوي الأخطاء". ونظرًا لأن الأخطاء قد تختلف، فإن متوسطات وتباينات البنود قد تختلف أيضًا. والافتراضات الأكثر تحرراً لهذا النموذج تكون جذابة لأن العثور على مقاييس متكافئــة للتباين المتســاوي هو أمر نادر الحدوث. ويتيح لنــا هذا النموذج أن نتوصل إلى العديــد من الاســتنتاجات ذاتها التي توصلنا إليها كما في الاختبــارات المتكافئة أو المتوازية التامة ولكن مع افتراضات أقل تقييدًا. وقد يرغب القراء في مقارنة هذا النموذج بمناقشــة «نونالي وبيرنستين» (Nunnally and Bernstein, 1994) لنموذج معاينة النطاق أو المجال.

يعتــبر بعض مطــوري المقاييس أن نموذج "تــاو" المترادف أو المتكافئ في الأســاس هو نمــوذج مقيد ومتشــدد جــداً. وفي نهايــة المطاف، كم مــرة يمكننا أن نفــترض أن كل بند يتأثــر بالمتغــير الكامن بنفــس الدرجة؟ الاختبــارات التي تم تطويرها في إطار ما يســمى بالنمــوذج المتقارب أو التقــاربي congeneric model (جوريســكوج، Jöreskog, 1971) تخضــع لمجموعه مــن الافتراضات الأكثر مرونة "غير المشــددة" (انظــر كارمينز، وماكلفير Carmines & McLver, 1981، لمناقشــه الاختبارات المتقاربــة أو التآلفية). يفترض هذا النموذج (فيما يتجاوز افتراضات القياس الأساســية) بكل بســاطة أن كافة البنود تشترك في متغــير كامن عام. وأنهــا لا تحتاج إلى أن يكون لها علاقات قويــة بنفس القدر مع المتغير

الكامــن، ولا ينبغي أن تكون تباينات الخطأ الخاصة بها متســاوية. ويجب أن يفترض المرء فقــط أن كل بند يعكس الدرجة الحقيقية إلى حــدٍ ما. وبطبيعة الحال، فإنه كلما كان كل بند يرتبط بقوة بالدرجة الحقيقية، كان المقياس أكثر ثباتاً.

المدخــل الأقل تقييدًا يتمثل في نموذج العامل العام general factor، الذي يتيح للمتغيرات الكامنة المتعددة أن تشــكل الأســاس لمجموعة معينة من البنود. وقد ناقش كل من "كارمينز وماكلفــير" (Carmines and McLver, 1981)، و"لوهلــين" (Loehlin,1998)، و"لونــج" (Long ,1983) مزايــا هــذا النوع مــن النماذج العامــة للغاية، وأهمها تحســين التطابق مع البيانات في العالم الحقيقي. وغالبًا ما تتضمن أساليب نمذجة المعادلة البنائية التحليل العاملي factor analyses في نمــاذج القيــاس الخاصــة بها؛ وتُعد الحالات التي تشــكل فيها المتغيرات الكامنة المتعددة مجموعة من المؤشرات مثالاً على نموذج العامل العام (لويهلن، Loehlin, 1998).

ويُعــد النموذج المتقارب أو التقاربي هو حالــة خاصة لنموذج العامل (أي حالة العامل المفــرد أو الوحيد). وبالمثل، فإن مقياس نموذج "تاو" المترادف أو المتكافئ في الأســاس هو حالــة خاصة من المقياس المتقــارب - الذي يفترض أن تكون العلاقات بــين البنود والمتغير الكامن الخاص بها متساوية. وأخيرًا، يمثل الاختبار المتكافئ التام حالة خاصة لنموذج «تاو» المترادف أو المتكافئ في الأســاس، مما يضيف افتراض وجود علاقات متســاوية بين كل بند ومصادر الخطأ المرتبطة به.

وهنا تجدر الإشارة إلى إستراتيجية قياس أخرى. هذه الإستراتيجية هي نظرية الاستجابة للبنــد أو المفردة («IRT» item response theory). تم اســتخدام هذا المدخل في المقام الأول، ولكن ليس بشكل حصري، مع بنود الاستجابة الثنائية (على سبيل المثال، الصحيح في مقابــل غير الصحيح) في تطوير اختبارات القدرة. وتفترض نظرية الاســتجابة للبند أن لكل بند من البنود على حدة حساســية خاصــة به للمتغير الكامن، ممثلــه بالمنحنيات المميزة للبند - وهو رسم بياني للعلاقة بين قيمة المتغير الكامن (على سبيل المثال، القدرة) واحتمال اســتجابة معينة لأحد البنود (على ســبيل المثال، الإجابة عليه بشــكل صحيــح). وبالتالي، يكشف المنحنى المميز للبند عن مقدار القدرة التي يتطلبها بند ما لتتم الإجابة عليه بشكل صحيح. سنتناول بشكل أكبر نظرية الاستجابة للبند في الفصل السابع.

فيما عدا دراسة نظرية الاستجابة للبند (IRT) في الفصل السابع ومناقشة التحليل العاملي في الفصل السادس، سوف نركز بشكل أساسي على النماذج المتكافئة أوالمتوازية ونموذج "تاو" المترادف أو المتكافئ في الأساس وذلك لعدة أسباب. أولاً، أنها تجسد وتمثل نظرية القياس «الكلاسيكية». ثانياً، أن مناقشة الآليات التي تعمل بها النماذج الأخرى بصورة سريعة يمكن أن تكون أمرًا مرهقًا. وأخيرًا، أثبتت النماذج الكلاسيكية أنها مفيدة جدًا للعلماء الاجتماعيين الذين لديهم اهتمامات أساسية غير القياس، ومع ذلك، فإنهم يأخذون القياس الدقيق على محمل الجد. وهذه المجموعة من الباحثين هي الجمهور الذي تمت كتابة هذا الكتاب له. وبالنسبة لهؤلاء الأفراد، فإن إجراءات تطوير المقياس التي تتبع النموذج الكلاسيكي تسفر عمومًا عن مقاييس مرضية. وفي الواقع، وعلى حد علمي أنه على الرغم من عدم وجود حصيلة متاحة بسهولة، أعتقد أنه (اختبار القدرة الخارجية) قد تم تطوير الغالبية العظمي من المقاييس المعروفة والمعترف بها بدرجة كبيرة والمستخدمة في أبحاث العلوم الاجتماعية باستخدام مثل هذه الإجراءات.

اختيار نموذج سببي:

إن اختيار النموذج السببي الذي يشكل أساساً لمتغير ما، عندما يكون ذلك ممكنًا، يمكن أن يكون جانبًا مهمًا من جوانب القياس. يمكن أحياناً تكييف مفهوم المتغير ذاته بدقة في بداية مشروع بحثي لجعل قياسه في النهاية أكثر سهولة وقابلية للتطبيق. على سبيل المثال، لنأخذ في الاعتبار أن باحثاً يريد تقييم كيفية تأثير بيئة العمل المادية على إنتاجية الموظف. قد يكون أحد الأساليب هو وضع قائمة طويلة من العوامل البيئية التي يُعتقد أنها تؤثر على الإنتاجية - مثل الإضاءة أو الإحساس بالخصوصية أو الوصول إلى جهاز حاسب آلي - وتطوير أداة تجعل العاملين يحددون مدى وجود هذه العوامل في مكان عمل معين. تتمثل مشكلة هذا المدخل أو الأسلوب في أن الأداة قد تكون في النهاية عبارة عن دليل index وليست مقياسًا أو ربما مزيجًا من الاثنين. أي إن المؤشرات (مثلاً، الإضاءة الجيدة والخصوصية المعقولة والوصول إلى جهاز حاسب آلي) قد لا تشترك في الواقع في مسألة مشتركة بل قد يكون لها تأثير مشترك، ألا وهو تحسين بيئة العمل. وبدلاً من ذلك، إذا نظر الباحث في مشكلة القياس المحتملة في وقت مبكر من عملية البحث، فإنه ربما يكون قد قرر تحديد

أو وضع مفهوم للمتغير بطريقة مختلفة إلى حد ما. فعلى سبيل المثال، إذا كان الباحث قد عـرّف المتغير محل الاهتمام على أنه تصورات الموظفين لبيئة العمل، فإن هذا التعريف قد يؤدي إلى مجموعة أكثر مرونة من البنود. على سـبيل المثال، يمكن أن يُطلب من الموظفين الموافقـة على بنود مثل: "بيئة مكان العمل الخاص بتوفر الأجهزة الأساسـية التي أحتاجها للقيام بعملي بشـكل فعال". هنا، لا يكون المتغير الكامن سـمه من سـمات البيئة في حد ذاتها ولكنه تصورات الموظفين. إن الطريقة التي يدرك بها الموظفون البيئة على أنها السبب العام الذي يقود ردودهم على كل بند من البنود على حدة، أجد أنه من الأسهل افتراض أن الموظف لديه شـعور ببيئة العمل التي سـتؤدي إلى إجابات عبر مجموعه من الأسئلة حول مدى كفايتها بدلاً من تصور البيئة نفسها كسبب لإجابات الموظف. وعلاوة على ذلك، فإن الطبيعة النفسـية لتصورات الموظفـين قد تكون في الواقع أقرب إلى مـا يعتبره الباحث ذا صله بالإنتاجية من مجرد وجود أو عدم وجود سـمات بيئية محدده. أي، إذا كان العامل ينظر إلى البيئة على أنها مواتية للإنتاجية قد يكون متغيراً أكثر أهميه من حكم شخص آخر فيـما يتعلق بمدى كفاية بيئة العمل. ولذلك؛ فإن وضع تصور للمتغير محل الاهتمام بهذه الطريقة قد يخدم السـؤال البحثي الأساسي بشكل جيد، مع احتمال تسهيل القياس النهائي للمتغير أيضاً.

وبطبيعة الحال، إذا لم يكن المتغير ذاته يقدم تصوراً سـببياً يتفق مع إستراتيجية القياس المباشرة، فلا ينبغي المسـاس بسـلامة المتغـير موضع الاهتمام. قد تسـاعد بعض المداخل والأسـاليب الباحث في التغلب على القيود المتأصلة في المتغير والطريقة التي يتم بها قياسه إجرائياً. ولكن إذا أمكن تبسيط تصور بديل مقبول للمتغير والنموذج الذي يربطه بمؤشراته، فقد يكون من الممكن تطوير أداة قياس تلبي وتفي بمجموعة أبسـط من الافتراضات، ومن ثم يمكن استكشافها باستخدام أدوات تحليلية أقل تعقيدًا. ومن المؤكد أن امتلاك الأدوات اللازمـة للتعامل مـع المواقف والحالات الأكثر تعقيـدًا يُعد أمراً جيـداً، ولكن تجنب تلك التعقيدات واستبعاد الحاجة إلى هذه الأدوات الأكثر تقدماً قد يكون أفضل، على افتراض أنه يحقق العدالة في المفهوم البنائي.

تمارين:

١- كيــف يمكننا اســتنتاج العلاقة بــين المتغير الكامــن والبندَين المرتبطَين به على أســاس الارتباطات بين البندَين؟

٢- مــا هو الفرق الرئيــسي في الافتراضات بين الاختبارات المتكافئــة ونموذج «تاو» المترادف أو المتكافئ في الأساس؟

٣- ما هو نموذج القياس الذي يفترض، إلى جانب الافتراضات الأساسية الشائعة لجميع طرق القياس، أن البنود تشترك في متغير كامن مشترك؟

٤- افترض أن نموذج «تاو» المترادف أو المتكافئ في الأســاس مع درجة حقيقية T ومؤشرات A وB وC. في مثــل هذا النمــوذج، يجب أن يكون لأي مؤشرين (على ســبيل المثال، A وB) والذين يشتركان في درجة حقيقية لهما تغاير مطابق للتغاير بين زوجين آخرين مــن المؤشرات (على ســبيل المثــال، B وC) يشــتركان في تلك الدرجــة الحقيقية. ومع ذلك، لا يجب أن تكون الارتباطات بين أزواج المؤشرات المختلفة متساوية. اشرح سبب حدوث ذلك.

ملاحظات:

١- على الرغم من أن (-٠,٧٠) هو أيضًا الجذر التربيعي المســموح به للقيمة (٠,٤٩)، إلا أن اتخاذ القرار بين الجذر الموجب أو الســالب عادةً ما يكون أقل أهمية مما يعتده المرء. طالما أن جميع البنود يمكن أن تكون مرتبطة بشــكل إيجابي مع بعضها البعض (إذا لزم الأمــر، من خلال عكس درجات بنود معينة، كما هو موضــح في الفصل الخامس)، فإن إشــارات معاملات المســار من المتغير الكامن لكل بند من البنود على حدة سوف تكون هي نفسها. لاحظ، مع ذلك، أن إعطاء إشارات إيجابية لهذه المسارات يعني ضمنياً أن البنود تشــير إلى المفهوم البنائي بشــكل أكبر، في حين أن المعاملات الســلبية تنطوي على عكس ذلك.

الفصل الثالث
الثبات (الموثوقية)

يُعد الثبات أو الموثوقية قضية أساسية في القياس النفسي. وأهمية الثبات تتضح بمجرد فهم معناه بشكل كامل وجيد. وكما يوحي المصطلح، فإن الأداة الموثوقة أو الثابتة هي الأداة التي تعمل بطرق متسقة ويمكن التنبؤ بها. ولكي يتمتع المقياس بالثبات، يجب أن تمثل الدرجات التي تنتج عن استخدامه بعضًا من الحالة الحقيقية للمتغير الذي يجري تقييمه وقياسه. ومن الناحية العملية، يعني هذا ضمنياً أن الدرجة التي تم الحصول عليها من استخدام الأداة يجب ألا تتغير ما لم يكن هناك تغيير فعلي في المتغير الذي تقيسه هذه الأداة، وبالتالي، يمكن أن يعزى أي تغيير ملحوظ في الدرجات إلى التغيير الفعلي في هذا المتغير. ومن شأن وضع مقياس ثابت تمامًا أن يكون انعكاسًا للدرجة الحقيقية وليس أي شيء آخر. ونادراً ما يمكن تحقيق ذلك؛ ومع ذلك، يمكننا أن نقيس مدى قربنا من هذه الحالة المثالية. وكلما كانت الدرجة التي نحصل عليها من مقياس تمثل الدرجة الحقيقية للمتغير وكلما كانت تعكس عوامل خارجية أخرى بدرجة منخفضة، كان المقياس الذي لدينا أكثر ثباتاً. وبشكل أكثر رسمية، فإن ثبات المقياس هو نسبة التباين التي تُعزى إلى الدرجة الحقيقية للمتغير الكامن. وهناك العديد من الطرق والأساليب لحساب الثبات، ولكنها تشترك جميعاً في هذا التعريف الأساسي.

وعلى الرغم من أن الطرق البديلة لحساب الثبات قد تبدو مختلفة، فإن التعريف الأساسي المشترك يتطلب أن تكون متكافئة حسابياً بطريقه أساسيه وهامه. وهذا هو الحال بالفعل. وتنطوي كل هذه الأساليب والطرق على تقدير الدرجة الحقيقية للمتغير وتحديد نسبة درجة المقياس التي تم الحصول عليها والتي تمثلها الدرجة الحقيقية. ويقترح نموذج القياس الأساسي، والذي تم توضيحه في الفصل الثاني، أن الدرجة الملاحظة تمثل مجموع الدرجة الحقيقية للمتغير الذي يجري تقييمه بالإضافة إلى الخطأ الناشئ عن عوامل خارجية. وتبعاً لذلك، نستطيع تقدير الدرجة الحقيقية للمتغير عن طريق طرح التباين الناتج عن الخطأ من التباين الكلي للدرجة الملاحظة التي تم الحصول عليها من مقياس معين.

ويمكننا بعد ذلك حساب الثبات كنسبة للدرجة الحقيقية المقدرة إلى الدرجة الملاحظة. وهكذا:

فإن الدرجة الحقيقية (true score) = الدرجة الملاحظة (observed score) – الخطأ (error)

الثبات = الدرجة الحقيقية / الدرجة الملاحظة

وتُعد طرق وأساليب تقدير الخطأ هي إلى حد كبير ما يميز المعادلات البديلة لحساب الثبات. وقد تم تصميم وتخصيص طرق مختلفة لأنواع محددة من البيانات، على الرغم من اشتراكها جميعاً في أساس مفاهيمي مشترك وهو: أن الثبات هو نسبة التباين في الدرجة الملاحظة التي يمكن أن تُعزى إلى الدرجة الحقيقية للمتغير الذي يجري تقييمه.

الطرق القائمة على تحليل التباين:

تعتمد إحدى وسائل تقدير الخطأ على تحليل التباين (ANOVA). ويقسم هذا المدخل التحليلي للبيانات التباين الكلي الملاحظ إلى مصادر مختلفة، لا سيما تلك التي تكون لها أهمية جوهرية (أي إشارة) وتلك التي تنشأ من مصدر خطأ ما (مثل الضوضاء)، مثل الخلل أو العيوب في أخذ عينة المشاركين من مجتمع ما. وعلى الرغم من أن هذا ليس هو المدخل الذي سنركز عليه في تقييم مدى ثبات المقاييس والقياس، فإن التطرق إليها بشكل موجز يؤكد على الاستمرارية عبر تعريفات ومداخل الثبات.

ومن ثم، من خلال استعراض ومراجعة سريعة، فإن النظر في مجموعة بسيطة للغاية من الملاحظات التي تنطوي على درجات حرارة ثمانية أجسام متماثلة، أربعة منها في ضوء الشمس المباشر وأربعة منها في الظل (لقد حددت عددًا صغيرًا من الأجسام الملاحظة في هذا المثال من أجل التبسيط)، والأجسام متشابهة باستثناء تعرضها لآشعة الشمس؛ ومع ذلك، فإن مقياس الحرارة المستخدم لقياس درجة حرارتها مشكوك فيه بعض الشيء، وبالتالي، فهو مصدر محتمل للخطأ في درجات الحرارة الملاحظة. ويمكننا تقييم مدى هذا الخطأ من خلال تسجيل درجات حرارة كل جسم من هذه الأجسام الثمانية على حدة وترتيب المعلومات بعدة طرق. أولاً، يمكننا تلخيص المعلومات حول الأجسام كمجموعة واحدة عن طريق حساب المجموع الكلي لمربع الانحرافات في درجات حرارة الجسم عن المتوسط الكلي لجميع الأجسام. ستكون هذه القيمة هي مجموع المربعات الكلي، أو SST.

ومن خلال تقسـيم مجموع المربعات الكلي على درجـات الحرية المرتبطة بالعينة بأكملها (على سـبيل المثال، N - 1 = 8 - 1 = 7)، سـوف نحصل على التباين الكلي لدرجات حرارة الأجسام. والخطوات التالية تعزل المكونات الفرعية لهذا التباين الكلي. ويمكننا المضي قدمًا في تقديـر مدى تأثير الخطأ على هذه الدرجات وبالتـالي المكون الفرعي من التباين الكلي. في إطـار تحليل التباين ANOVA، يتم تحقيق ذلك من خـلال تقييم مقدار التباين الذي يحدث في ظل ظروف مماثلة. في هذه الحالة، تتعرض جميع الأجسام الموجودة في الشمس لظـروف متماثلة أو متطابقة، وكذلك جميع الأجسـام الموجـودة في الظل. داخل كل من هاتين المجموعتين الفرعيتين، يُفترض أن تكون الأجسـام نفسها متطابقة وأن وجود أو عدم وجود ضوء الشـمس متماثل. لذلك؛ يجب أن يكون الأساس الوحيد للاختلافات في درجات الحرارة الملاحظة هو نوع من أنواع الخطأ. وبالتالي، يمكننا دراسة وفحص التباين في درجات حرارة الأجسـام داخل المجموعات لحسـاب مجموع مربعات الخطأ (SSE). وعن طريق طرح مجمع مربعات الخطأ SSE من مجموع المربعات الكلي SST، يمكننا حساب مجموع المربعات لتأثير ضوء الشمس. وهذا المجموع الأخير هو أساسًا مجموع المربعات للدرجات الحقيقيـة، أي إنه مؤشر على مقدار التباين في درجات حرارة الجسـم بعد إزالة تأثير خطأ القيـاس. ويمكننا بعد ذلك حسـاب تبايـن الدرجة الحقيقية من مجمـوع المربعات هذا. وأخيرًا، من خلال حسـاب النسـبة بين تباين الدرجات الحقيقيـة والتباين الكلي، نصل إلى نسـبة التباين الكلي التي يمكن أن تعزى إلى الدرجة الحقيقية (أي تأثير الشمس). ويمكننا تفسير هذه النسبة على أنها ثبات قياسنا لدرجة حرارة الأجسام.

لاحظ أنه إذا كانت جميع الأجسـام الموجودة في الشـمس لها درجـات حرارة متماثلة وكانت جميع الأجسـام الموجودة في الظل لها درجات حرارة متماثلة، يفترض أنها منخفضة، فإن تباين الخطأ سيسـاوي القيمـة "صفر". وبالتالي، لن يتم طـرح أي شيء من مجموع المربعات الكلي SST الملاحظة، وسـيكون تباين الدرجة الحقيقية والتباين الكلي متساويَين، وتكون النسبة التي تمثل ثبات قياس درجة حرارة الجسم تساوي الواحد الصحيح.

لقد أشرت إلى النسبة الناتجة عن مثال تحليل التباين الموضح أعلاه كمعامل ثبات، وهو صحيح. ومع ذلك، وبشـكل أعم، فإن النسـبة التي تقارن التباين الناتج عن بعض المصادر المحـددة في تصميم تحليل التباين مع التباين الكلي تُعرف باسـم معامـل الثبات المعياري أو معامل الارتباط داخل الطبقة (intraclass correlation coefficient)، أو اختصاراً ICC.

واعتــماداً على نــوع وتعقيد تصميم تحليل التباين، يمكن أن يكــون هناك أنواع عديدة من معامل الارتباط داخل الطبقة والتي ســتكون لها تفســيرات مختلفة، وليس كل منها تعادل أو تساوي ثبات القياس. على الرغم من أن القراء قد لا يكونون على دراية بمعامل الارتباط داخل الطبقة كما هو الحال بالنسبة للتعبيرات أو المصطلحات الأخرى الأكثر شيوعًا للثبات، إلا أننا سنرى أن المنطق الذي تستند إليه المزيد من المؤشرات الأكثر تخصصاً للثبات يتطابق مــع منطــق معامل الارتباط داخل الطبقة – وهذا يعنــي أن كلاً من معامل الارتباط داخل الطبقة والطرق الأخرى للحصول على الثبات تستند إلى مقارنة بعض تقديرات تباين الدرجة الحقيقية مع التباين الكلي.

البنود المتصلة «المستمرة» مقابل البنود ثنائية الاستجابة:

على الرغم من أن البنود قد تكون لها أشكال أو تنسيقات استجابة متنوعة، فإننا نفترض في هــذا الفصل أن الاســتجابات على البند تتألــف من خيارات اســتجابة متعددة القيمة. وتســتخدم البنود ثنائية الاستجابة (أي البنود التي لها خياران فقط للاستجابة، مثل «نعم» و«لا»، أو تلــك التي لديها خيارات اســتجابة متعددة يمكن تصنيفهــا على أنها «صحيحة» مقابل «خاطئة») على نطاق واسع في اختبار القدرة، وبدرجه أقل، في سياقات قياس أخرى. وتتضمن أمثلة البنود ثنائية الاستجابة ما يلي:

١- زيوريخ هي عاصمه سويسرا. ( ) صحيح ( ) خطأ

٢- ما هي قيمه «باي» pi؟

(أ) ١,٤١

(ب) ٣,١٤

(ج) ٢,٧٨

وقــد تم تطوير أســاليب وطرق خاصة لحســاب الثبات والتي تســتفيد من بســاطة العمليات الحسابية للاستجابات ثنائية الخيارات. وتغطي كتب القياس العامة مثل "نونالي وبرنشــتاين" (Nunnally and Bernstein, 1994) هــذه الأســاليب والطــرق بشيء من التفصيل. إن منطق هذه الأساليب لتقييم الثبات يماثل إلى حد كبير المدخل الأكثر عمومية

الذي ينطبق على البنود المتعددة الاستجابة والمقاييس المتصلة أو المستمرة. والواقع، إن المدخل المتبع في تقييم ثبات البنود متعددة الاستجابة هو في بعض الحالات امتداد لمدخل سابق تم وضعه وتطويره للبنود ذات الاستجابة الثنائية. وحرصاً على الإيجاز، فإن هذا الفصل لن يشير إلا بصورة عابرة إلى أساليب تقييم الثبات المعدة للقياسات التي تتكون من بنود ذات استجابات ثنائية. وتمت مناقشة بعض خصائص هذا النوع من المقاييس في الفصل الخامس.

الاتساق الداخلي:

إن ثبات الاتساق الداخلي Internal consistency reliability، كما يوحي الاسم، يتعلق بتجانس البنود داخل المقياس. وتهدف المقاييس القائمة على نماذج القياس الكلاسيكية إلى قياس ظاهرة واحدة. كما رأينا في الفصل السابق، تشير نظرية القياس إلى أن العلاقات بين البنود ترتبط منطقياً بعلاقات البنود بالمتغير الكامن. وإذا كانت لبنود مقياس ما علاقة قوية مع متغيرها الكامن، فستكون لها علاقة قوية مع بعضها البعض. وعلى الرغم من أننا لا نستطيع أن نلاحظ مباشرةً العلاقة بين البنود والمتغير الكامن، إلا أنه يمكننا بالتأكيد تحديد ما إذا كانت البنود مرتبطة ببعضها البعض. ويكون المقياس متسقًا داخليًا بقدر ما تكون بنوده مرتبطة ارتباطاً عالياً ببعضها البعض. ما الذي يمكن أن يفسر الارتباط بين البنود؟ هناك احتمالان: إما أن تؤثر البنود سببياً على بعضها البعض (على سبيل المثال، يسبب البند "أ" حدوث البند "ب")، أو أن البنود تشترك في قضية مشتركة. في معظم الحالات، يكون التفسير الأول غير مرجح، مما يجعل التفسير الأخير هو الخيار الأكثر وضوحاً. وهكذا، تشير الارتباطات العالية بين البنود وبعضها البعض إلى أن جميع البنود هي المقياس (أي، إنها مظاهر) لنفس الشيء. إذا وضعنا الافتراضات التي تمت مناقشتها في الفصل السابق (لا سيما الافتراض بأن البنود لا تتشارك في مصادر الخطأ)، يمكننا أيضًا استنتاج أن الارتباطات القوية بين البنود تنطوي على روابط قوية بين البنود والمتغير الكامن. ومن ثم، يجب أن يتكون المقياس الأحادي البعد أو بعد واحد لمقياس متعدد الأبعاد من مجموعة من البنود التي ترتبط جيدًا مع بعضها البعض. المقاييس المتعددة الأبعاد التي تقيس عدة ظواهر - على سبيل المثال، مقاييس موضع أو موقع التحكم أو السيطرة في الصحة

(مقاييس موقع الضبط الصحــي Health Locus of Control scales)[1] المتعددة الأبعاد (والســتون، والســتون، وديفيليس، Wallston, Wallston, &DeVellis, 1978) - هي في الواقع تجسد مجموعة من المقاييس المترابطة؛ وكل "بُعد" هو مقياس بحد ذاته.

معامل ألفا Alpha:

الاتساق الداخلي عادةً ما يتساوى (أو يعادل) مع معامل كرونباخ Cronbach (١٩٥١) ألفا (α). وســوف ندرس معامل ألفا بشيء من التفصيل لعدة أســباب. أولاً، أنه يستخدم على نطاق واســع كمقياس للثبات. وثانياً، قد يكــون ارتباطه بتعريف الثبات أقل وضوحاً مما هو الحال بالنســبة للمقاييس الأخرى للثبات (مثل أساليب النماذج البديلة) التي تمت مناقشــتها مؤخراً. وبالتالي، قد يبدو معامل «ألفا» أكثر غموضاً من أساليب حساب الثبات الأخــرى لأولئك الذين ليســوا على دراية بكيفيــة عمله الداخلي. وأخيراً، فإن استكشــاف المنطق الذي يقوم عليه حســاب معامل «ألفا» يوفر أساســاً سليماً لمقارنه كيفية استيعاب الأساليب الحسابية الأخرى لجوهر ما نعنيه ونقصده بالثبات.

معادلة كــودر - ريتشاردســون ٢٠ (Kuder-Richardson formula 20) أو KR-20، كما هو معروف بشــكل أكثر شــيوعاً - هي نســخة خاصه من معامل "ألفا" للبنود ثنائية الاستجابة (انظر، علي ســبيل المثال، نونالي وبرنشتاين، Nunnally & Bernstein, 1994). ويتســاوى كل من معادلة كودر - ريتشاردســون KR-20 مع معامــل "ألفا" عندما تكون البنــود المكونة لمقياس هي بنود ثنائية الاســتجابة. ومع ذلك، وكما ذكر آنفاً، ســنركز على الشكل الأكثر عمومية الذي ينطبق على البنود التي لديها خيارات استجابة متعددة.

يمكنــك التفكير في كل الاختلافات أو التباينــات في مجموعه من درجات البنود على أنه يحــدث بســبب أحد أمرين هــما: (أ) التباين الفعلي بين الأفراد في الظاهرة التي يقيســها

(١) يشير موضع الضبط والسيطرة بصفة عامة إلى المدى الذي يمكن من خلاله أن يتحكم الشخص أو يسيطر على الأحداث التي تؤثر فيه، وهو نوعان: داخلي ومعناه أن الشــخص يعتقد بأنه يمكن أن يتحكم في حياته، وخارجي أن الشــخص يعتقد بأن قرارته وحياته يتم التحكم فيها عن طريق الظروف البيئية المحيطة أو عن طريق الصدف. وموضع الضبط الصحي مفهوم يشــير إلى المدى الذي يُعد الأشخاص أنفسهم وفقه قادرين على التحكم بوضعهم الصحي أو يعتبرون أن وضعهم الصحي يتقرر بفعل تحكم الآخرين، أو تقرره عوامل الصدفة بشكل رئيسي (المترجمين).

المقيــاس (أي التباين الحقيقي في المتغير الكامــن) أو (ب) الخطأ. وهذا صحيح لأن نماذج القياس الكلاســيكية تعرف الظاهرة (على ســبيل المثال، رغبة المرضى في التحكم أو السيطرة على التفاعــلات مع الطبيب) كمصدر لكافة التباين المشــترك، ويعرفون الخطأ على أنه أي تباين متبقي، أو غير مشترك، في درجات المقياس (على سبيل المثال، بند واحد يكون له معنى مــزدوج غير المعنى المقصــود). وهناك طريقه أخرى للتفكير في هــذا الأمر وهو اعتبار أن التباين الكلي يشــتمل على وجود اثنين من المكونات هما: الإشارة (أي الفروق الحقيقية في رغبه المرضى في التحكم والسيطرة) والضوضاء (أي، الفروق في الدرجة الناتجة عن كل شيء ما عدا الفروق الحقيقية في الرغبة في التحكم أو السيطرة). إن حساب معامل «ألفا»، كما سنرى، يقســم التباين الكلي بين مجموعه البنود إلى مكونين هما الإشارة signal والضوضاء noise (الخطأ). نسبه التباين الكلي التي تمثل الإشارة تساوي معامل "ألفا". وهكذا، هناك طريقه أخرى للتفكير في معامل «ألفا» هو أنه يســاوي الواحد الصحيح مطروحاً منه تباين الخطأ أو، على العكس من ذلك، فإن تباين الخطأ يســاوي واحداً صحيحاً ناقص معامل ألفا (تباين الخطأ = ١- معامل ألفا).

مصفوفة التغاير:

يســاعدنا فحص مصفوفة التغاير covariance matrix لمجموعة من بنود المقياس على تحقيق فهم الاتســاق الداخلي بشــكل كامل. وتكشف مصفوفة التغاير لمجموعة من بنود المقياس عن معلومات مهمة حول المقياس ككل.

وتُعد مصفوفة التغاير هي شكل أكثر عمومية لمصفوفة الارتباط correlation matrix. في مصفوفة الارتباط، تكون البيانات معيارية standardized (أي معايرة وحدات القياس)، مــع تحديد التباينات بالواحد الصحيح. وفي مصفوفــة التغاير، تكون البيانات المدخلة غير معياريــة (وحدات القيــاس الأصلية). وبالتالي، فإنه يحتوي عــلى نفس المعلومات، كما في مصفوفة الارتباط، لكنها في شــكل غير معياري، والبنــود القطرية (المائلة) لمصفوفة التغاير هــي التباينــات - تباينات البنود مع نفســها - تمامًا مثلما تكون الوحــدات الموضوعة على طول القطر الرئيسي لمصفوفة الارتباط هي تباينات المتغيرات المعيارية إلى الواحد الصحيح، وأيضًــا ارتباطاتها مع نفســها. والقيم غــير القطرية هي التغايــرات covariances، والتي

تعــبر عن العلاقات بين أزواج من المتغيرات غير المعياريــة تماماً كما هو الحال في معاملات الارتبــاط المعيارية. وهكذا، من الناحية النظرية، تتكون مصفوفة التغاير من (أ) التباينات variances (على القطر) لكل متغير من المتغيرات على حدة، و(ب) التغايرات covariances (خارج القطر) التي تمثل العلاقة غير المعيارية بين أزواج المتغيرات.

وفي الجــدول رقــم ٣-١ يتم عــرض مصفوفة التغايــر النموذجية لثلاثــة متغيرات هي (X1 وX2 وX3).

جدول (٣-١): التباينات والتغايرات لثلاثة متغيرات

	X_1	X_2	X_3
X_1	Var_1	$Cov_{1,2}$	$Cov_{1,3}$
X_2	$Cov_{1,2}$	Var_2	$Cov_{2,3}$
X_3	$Cov_{1,3}$	$Cov_{2,3}$	Var_3

ويتم عرض نفس المصفوفة بشــكل أكثر توافقاً وتجانســاً باســتخدام الرمــوز المعتادة أو المألوفة للمصفوفات والتباينات والتغايرات على النحو التالي:

$$\begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 \end{bmatrix}$$

مصفوفات التغاير للمقاييس متعددة البنود:

دعونــا نركز اهتمامنا على خصائص مصفوفة التغاير لمجموعة من البنود التي تشــكل، عند إضافتها معًا، مقياسًــا. تحتوي مصفوفة التغاير الموضحة أعلاه على ثلاثة متغيرات هي X1، وX2، وX3. افــترض أن هــذه المتغيرات هي في الواقع درجــات لثلاثة بنود وأن هذه البنود هي (X1، وX2، وX3) عند إضافتها معًا تشكل مقياسًا سنسميه (Y). ما الذي يمكن أن تخبرنا به هذه المصفوفة عن علاقة كل بند من البنود على حدة بالمقياس ككل؟

تحتــوي مصفوفات التغاير على عدد من الخصائص المثيرة للاهتمام (على الأقل، الجيدة، والمفيــدة). مــن بين هذه الخصائص حقيقــة أن إضافة كل البنــود في المصفوفات معًا (أي جمع التباينات، التي تقع على طول القطر للمصفوفات، والتغايرات، التي تقع خارج القطر) تعطي قيمة مســاوية تمامًا لتباين المقياس ككل، بافتراض أن البنود متساوية الأوزان. لذلك؛ إذا أضفنــا كل البنــود إلى مصفوفات التغاير الرمزية، فســيكون المجمــوع الناتج هو تباين المقياس (Y). هذا مهم للغاية وجدير بالتكرار والذكر: التباين في مقياس (Y) والذي يتكون من أي عدد من البنود يساوي مجموع كافة القيم في مصفوفات التغاير الخاصة بهذه البنود، بافتراض تساوي وزن البند[١]. وبالتالي، فإن تباين المقياس (Y) المكون من ثلاثة بنود متساوية في الوزن (X1و X2 وX3) له العلاقة التالية بمصفوفات التغاير الخاصة بالبنود، حيث:

$$C = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{2,3} \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 \end{bmatrix}$$

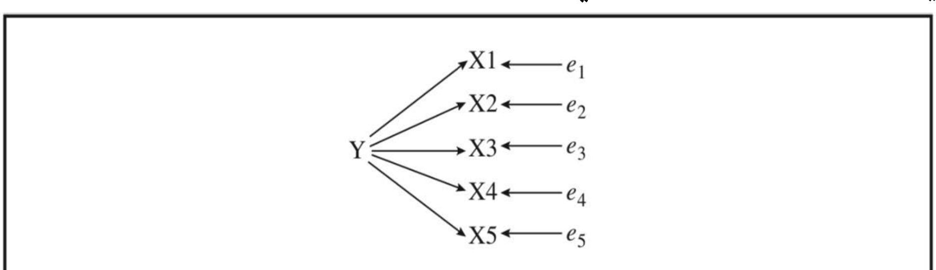
القــراء الذين يرغبون في الحصول على مزيد من المعلومــات حول الموضوعات التي تم تناولهــا في هذا القســم يمكن إحالتهم إلى نونــالي (Nunnally, 1978) من أجل مصفوفات التغايــر (التباين المشــترك) ونامبوديــري (Namboodiri, 1984) للحصول على مقدمة في مصفوفات الجبر في الإحصاء. تحتوي مصفوفات التغاير لكل بند من البنود على حدة على معلومات مفيدة أخرى. وقد ناقشت "بوهرنستد" (Bohrnstedt, 1969) التطبيقات التي يمكن استخلاصها من مصفوفات التغاير للبند.

معامل «ألفا» ومصفوفة التغاير:

تم تعريف معامل ألفا على أنه نســبة التباين الــكلي للمقياس التي يمكن أن تعزى إلى مصدر مشــترك، ومن المفترض أن تكــون الدرجة الحقيقية لمتغير كامــن تقوم عليه البنود. وبالتالي، إذا أردنا حســاب معامل «ألفا»، فســيكون من المفيــد الحصول على قيمة التباين الكلي للمقياس وقيمة لنســبة التباين «المشــترك». مصفوفات التغاير هي فقط ما نحتاج إليه من أجل القيام بذلك.

تذكــر المخطط البياني الذي اســتخدمناه في الفصل الثاني لتوضيــح كيفية ارتباط البنود بالمتغير الكامن الخاص بها، كما في الشكل رقم (٣-١).

جميــع الاختلافــات في البنود الناتجة عن المتغير الكامن (Y) هي مشــتركة أو شــائعة. (وتُستخدم مصطلحات المشتركة joint والشائعة communal أيضًا لوصف هذا الاختلاف). عندما يختلف المتغير Y (كما هو الحال، على سبيل المثال، بين الأفراد الذين لديهم مستويات مختلفة من الســمة التي يمثلها)، ســتختلف الدرجات على كافة البنود معه لأنه سبب تلك الدرجــات. لذلــك، إذا كانت درجة المتغير (Y) مرتفعة، فســوف تكون كل درجات البنود مرتفعــة؛ وإذا كانت درجة (Y) منخفضــة، فإن درجات البنود تميل إلى أن تكون منخفضة، وهذا يعني أن البنود ســوف تميل إلى التغيير بشــكل مشترك (أي، تكون مترابطة مع بعضها البعــض). لذلــك يؤثر المتغير الكامن عــلى كافة البنود، وبالتالي فــإن البنود تكون مرتبطة ببعضها البعض. وعلى النقيض من ذلك، فإن متغير الخطأ هو مصدر الاختلاف الفريد الذي يمتلكه كل بند. في حين أن كل البنود تشــترك في التباين بســبب (Y)، فلا يوجد بندان اثنان يشــتركان في أي تغيير من مصدر الخطأ نفسه في ظل افتراضات القياس الكلاسيكية الخاصة بنــا. تؤثــر قيمة متغير خطأ معين على درجة بند واحد فقــط. وبالتالي، لا ترتبط متغيرات الخطأ ببعضها البعض. ولذلك؛ فإن كل بند (وضمنيًا، المقياس الذي يحدده مجموع البنود) يتغــير كدالة لـ (أ) مصدر التباين المشــترك بينه وبين البنود الأخــرى، و(ب) التباين الفريد وغير المشــترك الذي نشــير إليه على أنه خطأ. ويترتب على ذلك أن التباين الكلي لكل بند - وبالتالي، بالنســبة للمقياس ككل - يجب أن يكون مزيجًا من التباين من المصادر المشــتركة والمصادر الفريدة. ووفقًا لتعريف الثبات، يجب أن تساوي قيمة معامل ألفا نسبة التباين في المصدر المشترك إلى التباين الكلي.



شكل (٣-١): مخطط يمثل كيفية ارتباط مجموعة مكونة من خمسة بنود بالمتغير الكامن المشترك (y)

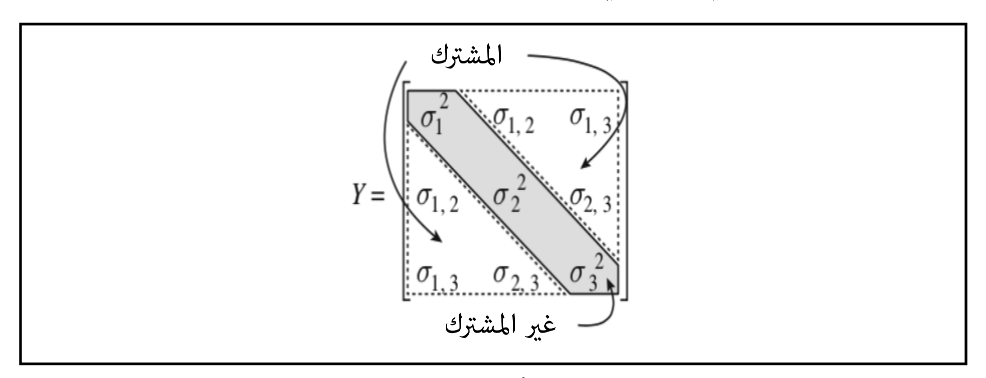
الآن، ضع في اعتبارك مقياساً من عدد (k) بند يسمى Y الذي تكون مصفوفة التغاير له على النحو التالي:

$$\begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \sigma_{1,3} & \cdots & \sigma_{1,k} \\ \sigma_{1,2} & \sigma_2^2 & \sigma_{1,3} & \cdots & \sigma_{2,k} \\ \sigma_{1,3} & \sigma_{2,3} & \sigma_3^2 & \cdots & \sigma_{3,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{1,k} & \sigma_{2,k} & \sigma_{3,k} & \cdots & \sigma_k^2 \end{bmatrix}$$

التباين (σ_Y^2) للمقياس المكون من عدد (k) بند يساوي مجموع جميع عناصر المصفوفة. المدخلات على طول القطر الرئيسي هي تباينات البنود الفردية الممثلة في المصفوفة. (ويُشار إلى تبايــن البند "i" كـ σ_i^2). لذلك؛ فإن مجمــوع البنود على طول القطر الرئيسي ($\Sigma\sigma_i^2$) هــو مجموع التباينات لكل بند من البنود على حــدة. وبالتالي، فإن مصفوفة التغاير تتيح لنا وصولاً ســهلاً إلى قيمتين هما: (١) التباين الــكلي للمقياس (σ_y^2) والذي يُعرف على أنه مجمــوع كل البنــود في المصفوفة و(٢) مجمــوع التباينات لكل بند مــن البنود على حدة ($\Sigma\sigma_i^2$) المحســوبة مــن جمــع القيم على طول القطــر الرئيسي. وهاتــان القيمتان يمكن إعطاؤهما تفســيرًا مفاهيمياً أو نظرياً. فمجموع المصفوفــة بأكملها، بحكم التعريف، هو تبايــن المتغير (Y)، للمقياس المكون مــن مجموعة بنود. ومع ذلك، فإن هذا التباين الكلي يمكن تقسيمه، كما قلنا، إلى أجزاء مختلفة.

دعنا ننظر في كيفية فصل مصفوفة التغاير المشتركة عن التباين الفريد من خلال دراسة كيفيــة اختــلاف البنود الموجودة في القطــر الرئيسي لمصفوفة التغاير عــن كافة البنود غير القطرية (أي تقع خارج القطر). كل التباينات (البنود القطرية) هي متغير مفرد أو "متغير مع نفسه". وقد لاحظنا في وقت سابق أن هذه التباينات يمكن اعتبارها بمثابة تغايرات في البنود مع نفسها. ويحتوي كل تباين على معلومات حول بندٍ واحدٍ فقط. وبعبارة أخرى، يمثل كل منها معلومات تســتند إلى بند واحد، وليس إلى التباين المرتبط المشــترك بين البنود.

(وفي إطـار هذا البند الواحد، سـيكون جزء من تباينه ناتجاً عن المتغير الأسـاسي المشـترك، وبالتالي، سيتم تقاسمه مع بنود أخرى؛ وبعضها لن يكون كذلك. ومع ذلك، فإن تباين البند لا يحدد مقدار التباين المشـترك ولكن فقط يحدد مقدار التشتت في الدرجات الخاصة بهذا البند، بغض النظر عن الأسباب). وتتضمن العناصر غير القطرية لمصفوفة التغاير أزواجاً من البنود، وبالتالي، التباين الشائع (أو المشترك) بين بندين من بنود المقياس (التغاير). وبالتالي، فـإن العناصر الموجودة في مصفوفة التغاير (وبالتالي التباين الكلي لـــلمتغير Y) تتكون من التغاير (التباين المشترك، إذا صح التعبير) بالإضافة إلى التباين "غير المرتبط" أو "غير المشترك" بشـأن البنود التي يتم النظر فيها بشـكل فردي. يمثل الشكل رقم (٣-٢) هذين التقسيمين الفرعيـين لمصفوفـة التغاير. والمنطقة المظللة على طول القطر هي الجزء غير المشـترك من المصفوفة، والمنطقتان غير القطريتين داخل الحدود المثلثية هما، معًا، الجزء المشترك.



شكل (٣-٢) مصفوفة التباين - التغاير تبين أن التباين على طول القطر الرئيسي هو التباين غير المشترك، في حين أن التغايرات الموجودة فوق القطر أو أسفله هي تباينات مشتركة

وبمـا أن التغايـرات - والتغايرات فقط - تمثل التباين المشـترك، فإنه يجب أن تكون كل التباينات غير المشـتركة ممثلة في التباينـات على امتداد القطر الرئيـسي لمصفوفة التغاير، وبالتـالي، يمكـن التعبير عنها بالمصطلـح ($\Sigma\sigma_i^2$). وبالطبع، يتم التعبـير عن التباين الكلي بـــ (S_y^2)، أي مجموع كافة عناصر المصفوفة. وبالتالي، يمكننا التعبير عن نسـبة التباين غير المشترك إلى التباين الكلي في المتغير (Y) على النحو التالي:

$$\Sigma\sigma_i^2 \ / \ \sigma_y^2$$

وتتطابق هذه النسـبة مع مجموع القيم القطرية في مصفوفة التغاير. وبالتالي، يمكننا أن نعبر عن نسبة التباين المرتبط أو المشترك، وما يتبقى منها – وبعبارة أخرى، تكملة هذه القيمة كما هو مبين كالتالي:

$$1- (\Sigma\sigma_i^2 / \sigma_y^2)$$

وتتوافـق هذه القيمة مع مجموع كافـة القيم غير القطرية لمصفوفة التغاير. قد يبدو غريباً، أو على الأقل غير فعال، حساب العناصر القطرية ثم طرحها من قيمة مصفوفة التغاير ككل. لماذا لا نحسـب فقـط مجموع العناصر غير القطرية بشـكل مباشر مثل ($\Sigma\sigma$i,j)، حيـث إن (i) و(j) يمثلان كل بند من البنديـن المتضمنين في تغاير معين؟ وفي الواقع، يمكن للمرء أن يصل إلى نفس النقطة بالضبط عن طريق الحسـاب المباشر لمجموع العناصر غير القطريـة. إن الصيغـة أو المعادلة التي تتضمن الطرح من الواحـد الصحيح هي إرث من الأيام التي لم تكن فيها أجهزة الحاسب الآلي متوفرة لإجراء العمليات الحسابية.

حساب التباين الكلي للمتغير (Y) والتباين لكل عنصر على حدة (i) ربما كانت عمليات تم تنفيذها بالفعل لأغراض أخرى. وحتى لو لم تكن هناك حاجة لحسـاب هذه التباينات لأغراض أخرى، فانظر إلى الجهد الحسـابي الذي تنطوي عليه. فبالنسـبة للمقياس المكون من ٢٠ بنداً، سـيكون الاختيار بين حساب ٢١ تبايناً (واحد لكل بند والآخر للمقياس ككل) مقابل ١٩٠ تغاير (أي واحد لكل عنصر من العناصر غير القطرية البالغ عددها ٣٨٠ عنصراً في المصفوفـة، مع تلك العناصر التي تكون فوق قطـر المصفوفة المتماثلة لتلك التي أدناه) بالإضافـة إلى التبايـن الكلي. ومن ثم، فإن الصيغة أو المعادلة التي تحدد التباين المشـترك على أنه ما تبقى بعد إزالة التباين غير المشترك من التباين الكلي، تكون معادلة أكثر منطقية مما قد تبدو عليه في البداية.

القيمة التي تمثلها المعادلة هي:

$$1- (\Sigma\sigma_i^2 / \sigma_y^2)$$

أو، أي ما يعادل أو يكافئ:

$$\Sigma\sigma_{i,j} / \sigma_y^2$$

مــن النظرة أو الوهلة الأولى قــد يبدو أن ذلك يعكس تعريف معامــل ألفا (أي الجزء المشترك من التباين الكلي في مقياس يمكن أن يعزى إلى المصدر المشترك للبنود، والذي نفترض أنــه يعكس الدرجة الحقيقية للمتغير الكامن). بيــد أننا ما زلنا بحاجة إلى تصحيح إضافي. وتصبــح هذه الحاجة واضحة إذا أخذنا في الاعتبار ما ســيحدث إذا كان لدينا، على ســبيل المثال، خمســة بنود مترابطة تمامًا. مثل هذا الترتيب ينبغي أن يؤدي إلى الثبات التام. إن مصفوفــة الارتباط، كما ذُكر آنفاً، هــي مجرد مصفوفة تغاير حيث تم فيها معايرة «توحيد قياس» كل البنود. ونظرًا لأن البنود الموجودة في مصفوفة الارتباط أكثر أُلفة واتســاقًا (أي أن الارتبــاط التام يتطابق دائمًــا مع معامل الارتباط الذي تكون قيمتــه الواحد الصحيح)، سأســتخدم مصفوفة الارتباط لتوضيح هذه النقطة. إذا استخدمنا التعبيرات المذكورة أعلاه لتمثيل جزء من التباين المشــترك والتباين الكلي في مصفوفة الارتباط الافتراضية الخاصة بها والمكونة من خمســة بنــود لها ثبات تام، فما الذي يمكن الحصــول عليه؟ تتكون مصفوفة الارتباط في هذه الحالة من مصفوفة من النوع (٥ x ٥) وكل القيم تساوي الواحد الصحيح. ومقام المعادلة السابقة، الذي يمثل التباين الكلي للمقياس الذي يتألف من البنود الخمسة، يســاوي ٢٥. ومع ذلك، فإن البســط لا يساوي سوى ٢٠ فقط، مما يسفر عن ثبات مقداره يســاوي ٢٥/٢٠ (أو٠,٨٠) بــدلاً من الواحــد الصحيح. لماذا يحدث هــذا؟ العدد الإجمالي للعناصر في مصفوفة التغاير هو (k^2). وعدد العناصر في المصفوفة التي تكون غير مشتركة (على ســبيل المثال، العناصر الموجودة على امتداد القطر الرئيسي) هو (k). والرقم المشترك (كل العنــاصر غير القطرية) هو ($k^2 - k$). وهكذا، فإن الكسر الموجود في المعادلة الأخيرة لدينا يحتوي على بســط يســتند إلى قيم ($k^2 - k$) والمقام يســتند إلى قيم (k^2). ولضبط أو تعديل حســاباتنا بحيث تعبر النســبة عن الأهمية النسبية بدلاً من عدد البنود المجمعة في البســط والمقام، فإننا نقوم بضرب المعادلة بأكملها التي تمثل نســبة التباين المشــترك في القيــم المقابلة للفروق في عدد البنود التــي تم تجميعها. وللقيام بذلك، نضرب المعادلة في المقــدار ($k^2 / (k^2 - k)$، أو، مــا يســاوي، ($k / (k - 1)$. وهذا الإجــراء يقيد مدى أو نطاق القيم المحتملة لــمعامل ألفا لتكون ما بين الصفــر والواحد الصحيح. في المثال المكون من خمســة بنود الذي تمت مناقشــته للتو، يــؤدي ضرب (٠,٨٠) في (٥ / ٤) إلى الحصول على القيمة واحد صحيح وهي قيمة الثبات التام. قد يرغب القراء في إجراء الحســاب الذهني

للمصفوفـات ذات الأحجام الأخرى. ينبغـي أن يصبح واضحًا أن المقدار $k / (k - 1)$ هي دائمًا المضاعف (المضروب فيه) الذي سـينتج عنه قيمة معامل ألفا مسـاوية لواحد صحيح عندما يكون الارتباط بين البنود وبعضها البعض تاماً. وهكذا، نصل إلى الصيغة أو المعادلة المعتادة لمعامل ألفا وهي:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\Sigma\sigma_i^2}{\sigma_y^2}\right)$$

وخلاصة القول، إن ثبات المقياس يسـاوي نسـبة التباين الكلي بين بنوده التي تُعزى إلى المتغير الكامن وبالتالي فهي مشـتركة. وتعبر صيغة أو معادلة معامل ألفا عن ذلك لتحديد الجزء من التباين الكلي لمجموعة البنود الفريدة، مع طرح هذا من الواحد الصحيح لتحديد النسـبة المشـتركة، والضرب في معامل تصحيح لضبط وتعديل عدد العناصر المسـاهمة في الحسابات السابقة.

صيغة أو معادلة بديلة لمعامل «ألفا»:

هنـاك معادلة شـائعة أخرى لحسـاب معامل ألفـا تعتمد على الارتباطـات بدلاً من التغايرات (التباينات المشـتركة). في الواقع، إن هذه المعادلة تستخدم متوسط الارتباط بين البنود ($\overline{r}$). هذه المعادلة هي:

$$\alpha = \frac{k\overline{r}}{1 + (k-1)\overline{r}}$$

وتتبع هذه المعادلة منطقياً للمعادلة القائمة على التغاير لمعامل ألفا. تأمل أو خذ بعين الاعتبار معادلة التغاير من حيث المصطلحات المفاهيمية:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\text{مجموع تباينات البند}}{\text{مجموع التباينات والتغايرات}}\right)$$

لاحظ أن البســط والمقــام في الحد على اليمين هما عبارة عن مجمــوع القيم لكل بند. ومــع ذلك، فإن مجموع هــذه القيم الفردية يكون مماثلاً لمتوســط القيم مضروباً في عدد القيم المعنية. (على سبيل المثال، أرقام (k) التي يساوي مجموعها٥٠ و(k) ضعف متوسط هذه الأرقام تساوي ٥٠ أيضاً. ولمزيد من التوضيح، بالتعويض عن k بالقيمة ١٠ في الإجراء السابق؛ يجب أن يكون متوسط ١٠ قيم يصل مجموعها إلى ٥٠ هو خمسة أضعاف القيمة عــشرة أو عشرة أضعاف القيمة خمســة، وهي نفس القيمة في المجمــوع الأصلي). لذلك؛ يجب أن يساوي البسط الخاص بالحد على اليمين (k) مضروباً في متوسط تباين البند ($\overline{v}$) ويجب أن يســاوي المقام (k) مضروباً في متوسط التباين زائد (k^2 - k) - أو، بدلاً من ذلك، (k - 1) (k) – مضروباً في متوسط التغاير ($\overline{c}$):

$$\alpha = \frac{k}{k-1}\left(1 - \frac{k\overline{v}}{k\overline{v} + (k)(k-1)\overline{c}}\right)$$

لإزالة الواحد الصحيح من المعادلة، يمكننا استبدالها بما يعادلها، / $\overline{c}$ (k–1) (k)+$\overline{v}$ k] [k $\overline{v}$+(k)(k–1) $\overline{c}$] مما يسمح لنا بـدمج الحد على اليمين بأكمله في نسبة واحدة:

$$\alpha = \frac{k}{k-1}\left(\frac{k\overline{v} + (k-1)\overline{c} - k\overline{v}}{k\overline{v} + (k)(k-1)\overline{c}}\right)$$

أي ما يعادل أو يكافئ:

$$\alpha = \frac{k}{k-1}\left(\frac{k(k-1)\overline{c}}{k\left[\overline{v} + (k-1)\overline{c}\right]}\right)$$

يؤدي الحذف المتبادل لـلمقدار (k) من البسط الخاص بالحد الأيسر ومقام الحد الأيمن، بينما يؤدي الحذف المتبادل للمقدار (k - 1) من البســط الخاص بالحد الأيمن ومقام الحد الأيسر إلى الحصول على الشكل المبسط التالي للمعادلة:

$$\alpha = \frac{k\overline{c}}{\overline{v} + (k-1)\overline{c}}$$

ونذكر أن الصيغة أو المعادلة التي نسعى للحصول عليها تنطوي على ارتباطات بدلاً من التغايـرات، وبالتالي متغيرات معيارية بدلاً مــن متغيرات غير معيارية. بعد المعايرة، يكون متوسط التغايرات مطابقاً لمتوســط الارتباطات ويساوي التباين الواحد الصحيح. وبالتالي، يمكننا استبدال متوسط الارتباط بين البنود وبعضها البعض ($\overline{r}$) مع الواحد الصحيح. وينتج عن ذلك الصيغة أو المعادلة القائمة على الارتباط لمعامل ألفا:

$$\alpha = \frac{k\overline{r}}{1 + (k-1)\overline{r}}$$

وتُعرف هذه الصيغة أو المعادلة باسم معادلة تنبؤ سبيرمان - براون Spearman-Brown، وسيتم توضيح أحد استخداماتها المهمة في قسم من هذا الفصل الذي يتناول حساب الثبات بطريقة التجزئة النصفية. وتحســباً لهذه المناقشــة، لاحظ أن المعادلة تقدم الثبات كدالة لخاصيتــين من خصائــص المقياس هما: عدد البنود في المقياس (k) ومتوســط الارتباط بين أزواج البنــود ($\overline{r}$). وبالتــالي، كلما زاد عــدد البنود التي يحتويها المقياس وكان متوســط الارتباط بين تلك البنود أقوى، زاد الثبات.

ويشــار أحيانًا إلى الصيغتين المختلفتين، التي تســتند إحداهما على التغايرات وتســتند الأخــرى عــلى الارتباطات، على أنهــما معادلات وصيــغ الدرجة الخام والدرجــة المعيارية لمعامل ألفا، على الترتيب. وتحافظ معادلة الدرجة الخام على المعلومات حول متوســطات وتباينات البند في عملية الحســاب لأن التباينات المشــتركة (التغاير) تســتند إلى القيم التي تحتفظ بوحدات القياس الأصلية للبيانات الخام. إذا كانت للبنود تباينات مختلفة بشــكل ملحوظ، فإن تلك البنود التي لها تباينات أكبر سيكون لها وزن أكبر من تلك البنود التي لها تباينات أقل عند استخدام هذه المعادلة لحساب معامل ألفا. ولا تحتفظ المعادلة المعيارية أو القياســية القائمة على الارتباطات بوحــدات القياس الأصلية للبنــود. تذكر أن الارتباط هــو تغايــر معياري أو قياسي. لذلك يتم وضع جميع البنود على مقياس مشــترك، وبالتالي، تكون الأوزان متســاوية في حســاب معامل ألفا من خلال المعادلــة المعيارية. أيهما أفضل

أن يتم الاعتماد على السـياق المحدد أم الوزن المتسـاوي؟ وكما سنرى في الفصول اللاحقة، فإن الإجراءات الموصى بها لتطوير البنود غالبًا ما تسـتلزم هيكلة صياغتها بحيث تسفر عن تباينـات مماثلة لكل بنـد. وعندما يتم اتباع هذه الإجراءات، يكـون هناك عادةً اختلاف بسيط في معاملات ألفا المحسـوبة بواسطة الطريقتين البديلتين. ومن ناحية أخرى، عندما لا يتـم اتبـاع الإجراءات الرامية إلى إحداث تباينات متكافئـة في البنود، يلاحظ أن قيم ألفا المعيارية والخام تختلف اختلافًا كبيرًا (على سبيل المثال، (٠,٠٥) أو أكثر) يدل على وجود بند واحد على الأقل له تباين يختلف اختلافًا كبيرًا عن التباينات في البنود الأخرى.

نقد معامل «ألفا»:

يمكـن للمرء أن يجادل بـأن معامل ألفاكرونباخ ليس مؤشراً مثالياً للثبات (على سـبيل المثال، سيجتزما Sijtsma, 2009). وفي الواقع، فإنه - أي معامل ألفاكرونباخ - عادةً ما يوفر الحـد الأدنى لتقديـر الثبات lower bound الحقيقي لمجموعة مـن البنود بدلاً من أفضل تقدير للثبات الحقيقي. ويشير «سيجتزما» (٢٠٠٩) إلى أن مفهوم الاتساق الداخلي غامض إلى حد ما ويثير تساؤلاً حول مدى تمثيل معامل ألفا لأهم جوانب الاتساق الداخلي. ويشير إلى أن الاتسـاق الداخلي يتعلق بالبنية العاملية لمجموعه من البنود وأن اعتبار معامل ألفا مؤشراً على ذلك هو إفراط في التبسـيط. بالمعني الدقيق أو الضيق، فإن رأي «سـيجتزما» صحيح. غير أنه يعترف بوجود مشاكل في التقديرات الأخرى (على سبيل المثال، تقدير الحد الأدنى المنخفض للثبات، greatest-lower-bound estimate of reliability, or glb) التي يمكن اعتبارها بديلاً عن معامل ألفا. ويقترح «سيجتزما»، بشكل مثير ومناسب، كسبيل إلى المزيـد من الدقة، ينبغي القيام بمزيد من العمل من أجل الفهم الكامل لأنواع التحيز التي قد تكون موجودة في مختلف مؤشرات الثبات.

هناك نقد آخر لمعامل ألفا كمعامل للثبات يتمثل في أنه تم تصميمه بشكل أساسي للبيانات المسـتمرة أو المتصلة. وتعتمد الحسابات التي تنطوي عليها عملية تحديد معامل ألفا على ارتباطات بيرسون Pearson، التي تفترض القياس المسافي أو الفتري interval scaling. ويلاحظ كل من "جادرمان، وجوهن، وذيمبـو" (Gadermann, Guhn, and Zumbo, 2012) أن أنواع الاسـتجابات أو الردود على البنود كثيراً ما تُسـتخدم في العلوم الاجتماعية وفي غيرها

هي، على وجه الدقة، من نوع القياس الترتيبي وليســت من نوع القياس المسافي أو الفتري. وكمثــال، خذ بعين الاعتبار البند الافتراضي التالي (وخيارات الاســتجابة المصاحبة له) والذي يقصد به أن يكون جزءاً من مقياس الشك الذاتي:

لم أتمكن من إنجاز وتحقيق كل ما كنت أتمنى أن أفعله.

١	٢	٣	٤	٥	٦
لا أوافق بشدة	لا أوافق بدرجة متوسطة	لا أوافق بدرجة قليلة	أوافق بدرجة قليلة	أوافق بدرجة متوسطة	أوافق بشدة

من الواضح أن الفئات مرتبة من الأقل اتفاقاً إلى الأكثر اتفاقاً مع البند، ولكننا لا نستطيع أن نفترض أن الزيادة في الشــك الذاتي التي يمثلها الفرق بين اختيار الاســتجابة (٣) بدلاً من الاســتجابة (٢) هو نفس مقدار الزيادة في الشــك الذاتي عند اختيار الاستجابة (٦) بدلاً من الاســتجابة (٥). أي، إنــه لا يمكننا أن نفترض زيادات المســافات بالتســاوي في المتغير الذي يســتنبطه البند. فالفئات مجرد فئات ترتيبية بالمعني الدقيق للكلمة. وقد أشار «جادرمان وآخرون» (Gadermann et al., 2012) إلى أن هذه المسافات غير المتساوية تؤدي إلى عدم الدقة في اســتخدام معامل ألفا كتقدير لثبات الاتســاق الداخلي وأوصوا ببديل آخر، أطلقوا عليــه معامل ألفا الترتيبــي **ordinal alpha coefficient**، والــذي لا يفترض القياس الفتري أو المســافي. ويستبدل مدخلهم أو طريقتهم هذه اســتخدام ارتباطات بيرسون بالارتباطات الرباعية **tetrachoric** أو متعددة الأقسام **polychoric** (الأولى في الحالات التي تنطوي على الاســتجابات الثنائية للبند، والأخيرة في الحالات التي تنطوي على خيارات اســتجابة ترتيبية متعددة النقاط). وتفترض الارتباطات الرباعية / متعددة الأقســام أن المتغير قيد الدراســة هو متغير مســتمر أو متصل لكن يتم قياسه بشــكل ترتيبي، باستخدام فئتين أو أكثر مرتبة كخيارات اســتجابة. ومن الناحية النظرية، يمكن للمرء أن يفكر في حدود الســمة المقاســة (الشك الذاتي في هذا المثال) الموجودة على طول سلسلة متصلة. عند مستوى ما من السمة، هناك حد، عندما يتم تجاوزه، يميل المجيب لاختيار خيار الاستجابة (٢) بدلاً من الخيار (١).

توجد حدود إضافية (عند مسافات غير محددة) بين كل خيار استجابة والبديل الأعلى التالي. إن الارتباط متعدد الأقسام بين أي بندين من النوع الذي تم وصفه سابقًا يبين أساساً مدى ميل هذه الانتقالات والتحولات عبر هذه الحدود إلى الحدوث بشكل مرتبط بدلاً من أن تحدث بشكل مستقل بالنسبة للبنود المعنية. هذا الارتباط هو تقدير الاحتمال الأقصى للارتباط الحقيقي بين المتغيرات الترتيبية (تيمرمان، ولورينـزو Timmermann & Lorenzo-Seva, 2011). باستبدال مصفوفة ارتباطات بيرسون بمصفوفة الارتباطات متعددة الأقسام، يمكن للمرء الحصول على معامل ألفا الترتيبي بدلاً من معامل ألفا المعياري.

وقام «غادرمان وآخرون» (Gadermann et al., 2012) ببحث تأثير استبدال ارتباطات بيرسون بالارتباطات الرباعية / متعددة الأقسام في دراسة تتضمن بيانات محاكاة simulated data. أظهرت نتائج دراسة المحاكاة هذه أن معامل ألفا يظهر تحيزًا أكبر (أي انخفاض التقديرات لأقصى درجة للثبات المعروف لمجموعة البنود المعينة) مقارنةً بمعامل ألفا الترتيبي. إن الالتواء الشديد في توزيع الدرجات يضر على وجه الخصوص بدقة التقديرات الناتجةعن معامل ألفاكرونباخ، وفقًا للمحاكاة التي أوردها «غادرمان وآخرون»، ومع ذلك، تجدر الإشارة إلى أن الارتباطات متعددة الأقسام تنطوي على قيود خاصة بها، لا سيما مع أحجام العينات الصغيرة (Timmermann & Lorenzo-Seva, 2011).

انتقد مؤلفون آخرون معامل ألفا لأسباب أخرى غير تحيزها عندما يكون قياس البند من النوع الترتيبي وليس من النوع الفتري أو المسافي. على سبيل المثال، لاحظ كل من «ديون، باجولي، وبرينسدن» (Dunn, Baguley, and Brunsden, 2014: p. 404) ما يلي:

الصعوبات الرئيسة في استخدام معامل ألفا كمقياس للاتساق الداخلي يمكن تلخيصها على النحو التالي:

١- يعتمد معامل ألفا على افتراضات بالكاد يمكن أن تتحقق (أي يصعب تحقيقها).

٢- يؤدي انتهاك هذه الافتراضات إلى تضخيم معامل ألفا وتقليل تقديرات الاتساق الداخلي للمقياس.

٣- "ألفا إذا تم حذف بند" [طريقة عامة لتقييم تأثير كل بند من البنود على حدة على معامل ألفا الكلي] في عينة لا يعكس تأثير حذف البند على ثبات المجتمع الإحصائي population.

٤- لا يعكــس التقدير بنقطة لمعامل ألفا التباين الموجــود في عملية التقدير، مما يوفر ثقة زائفة في اتساق إدارة وتطبيق مقياس ما.

ســأتناول كلاً من هذه الانتقادات تباعاً واحدة بعــد الأخرى. تتطلب افتراضات نموذج «تــاو» المتماثل أو المترادف في الأســاس التي يقــوم عليها معامل ألفــا أن تكون كل البنود مــؤشرات جيدة بنفس القدر لمفهوم بنائي واحد يشــتركون فيه. أي، إنه يجب أن يقوم أي بند بمهمة جيدة في التأثير على الســمة التي يتم قياســها مثل أي بند آخر. وبشــكل أكثر رســمية، يحــدد هذا الافتراض أن مجموعــة البنود تكون أحادية البعــد وأن التغايرات بين البنــود والدرجة الحقيقية للمتغير الكامن متســاويان أساســاً، على الرغــم من أن تباينات الخطأ (وبالتالي التباينات الملاحظة للبنود) لا يلزم أن تكون كذلك. ويمكن أن يحدث انتهاك أو مخالفــة الافــتراض الخاص بأحادية البعــد إذا كان محتوى البنــود ينطوي على أكثر من مفهوم واحد متميز تشــترك فيه البنود. على ســبيل المثال، قد يؤدي استخدام أداة تهدف إلى قيــاس القلــق إلى قيــاس جوانب الاكتئاب عن غــير قصد. وإذا لم يتــم إدراك وتحديد هذه الازدواجية المفاهيمية، فإن ذلك قد يشــكل انتهاكًا لافتراض أحادية البعد الذي يقوم عليه اســتخدام معامل ألفــا. وبطبيعة الحال، فإن افتراض أحاديــة البعد هو أيضًا افتراض لمعظم البدائل المقترحة لمعامل ألفا، وبالتالي، فإن انتهاك افتراض أحادية البعد يطرح مشكلة أكــثر عمومية في تقدير الثبات. وهناك نوع أكثر شــيوعًا مــن انتهاك نموذج «تاو» المتماثل أو المترادف في الأســاس هو وجود بنود لا تعبر عن المتغير الأساسي بنفس القدر من الصحة، أي أن لها تغايرات مختلفة مع الدرجة الحقيقية. وسيتضح ذلك في شكل ارتباطات متبادلة غير متجانســة بــين البنود. لأن كل التباين المشــترك بين بندين في مجموعــة بنود أحادية البعــد، هو مظهر مــن مظاهر ارتباط كل بند بالدرجة الحقيقيــة، وإذا كانت بعض البنود ارتباطها أقل قوة بتلك الدرجة الحقيقية من غيرها، فإن الارتباطات بين البنود التي تنطوي على تلك البنود ســتكون أقل. لذلك؛ بالنســبة لأربعة عناصر افتراضية هي A وB وC وD لهــا تغايرات خاصة بــكل منها مع الدرجات الحقيقية هــي ٠,٣١و٠,٣٠و٠,٢٧و٠,٠٥، فإن الارتباطات بين A وB وC (أي rAB وrAC، وrBC) بشكل عام ستكون أعلى (وأكثر تشابهًا) من الارتباطات التي تتضمن البندD (أي rAD وrBD وrCD). سيكون عدم تجانس تباين البنــود بمثابة دليلٍ مهمٍ عــلى انتهاك افتراض نموذج "تاو" المتماثل أو المترادف في الأســاس

(أي، مقارنــات التغايرات من البند إلى الدرجــة الحقيقية). كما أن البنود التي لها خيارات اســتجابة مختلفة، مثل خيار مكون من (٤) نقاط لبعض البنود وخيار مكون من (٧) نقاط للبنود الأخرى، كثيراً ما تؤدي أيضًا إلى انتهاك افتراض تماثل تغايرات البند والدرجة الحقيقية. لماذا هو كذلك؟ وحتى عندما يكون هناك شــخصان لهما نفس الدرجة الدقيقة من السمة التــي يجري تقييمها، فإن التطابق بين درجاتها على بندين لا يمكن أن تكون تامة إلا عندما يكون عدد خيارات الاســتجابة متطابقًا لكلا البندين. وبشــكل عام، فإن عدم تطابق عدد خيارات الاستجابة بين بندين سيسفر عن انخفاض درجة ارتباطهما معاً، مقارنةً بأداء نفس البنود عندما يكون لكل منهما نفس العدد من خيارات الاســتجابة. لذلك؛ ستؤدي خيارات الاستجابة غير المتطابقة إلى بعض أزواج البنود (تلك التي لها نفس عدد خيارات الاستجابة) التي من المحتمل أن يكون لها ارتباطات أقوى مع أزواج بنود أخرى (حيث لا يتطابق عدد خيــارات الاســتجابة). وبما أن الارتباطات بين البنود وبعضهــا البعض متأصلة في ارتباطات البنــود بالدرجة الحقيقية، فإن هذه الارتباطات المتباينة بــين البنود وبعضها البعض تعني ضمنيــاً أن التغايــرات بين البنود والدرجــة الحقيقة تكون مختلفة أيضًــا عبر البنود. هذا، بطبيعة الحال، يشــكل انتهاكًا لنموذج «تاو» المتماثل أو المترادف في الأســاس والذي يستند إليــه معامل ألفا. ونتيجة لانتهاك هذا الافتراض المحدد، فإن معامل ألفا قد لا يقدم تقديرًا دقيقًــا للثبات الحقيقي للمقياس. وهكذا، يجادل دان وآخرون (Dunn et al., 2014) بأن استمرار استخدام ألفا يطرح (أو يسبب) مشكلة.

أمــا حجتهم الثانية هي أن الفشــل في تلبية تلك الافتراضات والوفاء بها ليســت مجرد مســألة نظرية بل هي أيضاً نتيجة عمليه. أي إن أنواع الثبات تقدر التباينات التي ينتجها معامل ألفا كبيرة بما فيه الكفاية. على الرغم من أن كلاً من المغالاة في التقديرات والتقليل مــن التقديرات هما أمــران ممكنان عندما تفشــل مجموعه البنود في الوفــاء بافتراضات نموذج «تاو» المتماثل أو المترادف في الأســاس، فإن التقدير المنخفض هو الأكثر شيوعاً بكثير، لأن معامــل ألفا هو تقديــر الحد الأدنى للثبات في إطار نموذج «تــاو» المتماثل أو المترادف في الأســاس. ومع ذلــك، يمكن أن يحدث التضخــم عندما يكون عدد البنود التي تشــكل المقياس كبيراً جداً (لأن القيمة التي تم الحصول عليها لمعامل ألفا تميل إلى الزيادة مع عدد البنود). أما الظروف التي تنطوي على أكبر التناقضات بين مســتويات ألفا الاسمية والثبات

الحقيقي هي تلك التي تشكل انتهاكات صارخة للافتراضات المتعلقة لنموذج «تاو» المتماثل أو المترادف في الأســاس. وقد تم وصف بعض هذه الحالات في وقت ســابق عند «غادرمان وآخرين» (Gadermann et al., 2012).

أيضًــا، انتقد «دان» وآخــرون (Dunn et al., 2014) الممارســة الخاصة بحذف البنود بالتتابع واحدًا تلو الآخر، ثم النظر في التأثير على معامل ألفا كوسيلة لاكتشاف أفضل السبل لتحســين الاختصار للمقياس وثباته (موضوع سأناقشه بمزيد من التفصيل في فصل لاحق). وتؤكد الحجة التي طرحوها على أن حذف بند معين قد يزيد معامل ألفا لمجموعة متنوعة من الأسباب، وبالتالي قد لا تمثل زيادة في تغاير الدرجة الحقيقية للبند ككل. ويقترحون أن الأساس المنطقي لحذف بند ما يستند على افتراض أن جميع البنود لها تباينات خطأ متساوية (وهي خاصية أو ســمة من سمات البنود التي تفي أو تلبي افتراضات الاختبارات المتوازية أو المتكافئة ولكن ليســت أساسًا للاختبارات الخاصة لنموذج "تاو" المتماثل أو المترادف في الأســاس، على النحو المشــار إليه في الفصل الثاني). ولكن هذا لا يفترض في معظم تطبيقات معامــل ألفــا. وغالباً، يكون الافــتراض هو لنموذج «تاو» المتماثل أو المترادف في الأســاس، والذي لا يتطلب تباينات خطأ متساوية. وبالتالي، قد لا يكون للبند المحذوف تباين درجة حقيقية أكبر، ولكن يملك تباين خطأ أقل فقط عند استخدامه في هذه الحالة المحددة (أي، مــع العينة التي تم جمع البيانات منها). ومن المفــترض أن التقلبات أو التغيرات في تباين الخطــأ تعتمد بشــكل أكبر على العينــة، في حين أن التغاير في الدرجــة الحقيقية أكثر صلة بتقديــر مدى ثبات المقياس في المجتمع الإحصــائي. ومن ثم، فإن الحجة، تذهب إلى القول بأن التحسن الظاهر في الثبات الذي تحقق عن طريق إسقاط أو حذف البنود التي لا تعزز معامل ألفا قد يكون مجرد وهم أو خداع.

ويســتند انتقادهم النهائي لمعامل ألفا على فكرة أن تقدير قيمة وحيدة للثبات ليســت ذا مغــزى محدد ما لم يكــن لدينا بعض الفهــم والإدراك بمدى التشــتت الكبير لمثل هذه التقديــرات. أي إنــه بدون فترة ثقة حول تقدير الثبات الذي يوفره معامل ألفا، فإنه يكون من الصعب قياس مقدار الثقة التي ينبغي أن تكون لدينا في ذلك التقدير.

معالجة محددات وقيود معامل «ألفا»:

إحدى وســائل التغلب على هذا الانتقاد الأخير هي تحديد فترة ثقة لمعامل ألفا. وهذا يســمح للباحث بأن يتأكد أنه ضمن هامش خطأ محدد مســبقًا (على ســبيل المثال، الثقة بنسبة ٩٥٪)، يقع الثبات الحقيقي للمقياس بين قيمتين محددتين. لذلك؛ على سبيل المثال، إذا كانــت قيمــة معامل ألفا تبلغ ٠,٨٥ (تقدير النقطة للثبــات) لها فترة ثقة ٩٥ ٪ يحدها ٠,٨٠و٠,٩٠، فإن ذلك يشير إلى أن احتمال أن يقع الثبات الحقيقي داخل تلك الحدود كان بنسبة ٩٥ ٪ واحتمال أن يكون خارج تلك الحدود كان بنسبة ٥ ٪.

الطريقــة التي أوصى بهــا «دان وآخــرون» (Dunn et al., 2014) وغيرهم (على ســبيل المثال، باديــلا، وديفرز، ونيوتون Padilla, Divers, & Newton, 2013) لتحديد فترات الثقة لتقديــر الثبات المعين الــذي يتضمن تقنية[٢] bootstrapping. ولتوضيح تقنية بوتســترابنج bootstrapping، أقــوم أولاً بمراجعة كيفية إنشــاء فــترة ثقة بدون بوتســترابنج. وعادةً ما يتطلــب مفهوم فترة الثقة، كما ينطبق هنا، اســتخدام مجموعات بيانــات متعددة من أجل الحصــول على توزيع قيم لمعلمة المجتمع parameter التي يتم تقديرها. لذلك؛ على ســبيل المثال، إذا أردنا فترة ثقة حول تقدير نقطة لمتوســط طول للأشخاص في مدينة معينة، فيمكننا ســحب أو أخذ العديد من عينات الأشــخاص من تلك المدينة، وحســاب متوسط الطول لكل عينة، وإنشــاء توزيع لهذه المتوســطات. ويكون المتوسط الكبير (أي القيمة المتوسطة لتوزيع متوســطات العينة) هو تقديرنا لنقطة المتوسط الحقيقي لطول الأشخاص في المدينة. ونتوقع حــدوث اختلافــات في متوســط الأطــوال الملاحظة من عينة إلى أخرى بشــكل عشــوائي لأن جميــع العينات مأخوذة مــن نفس المجتمع. وبالتــالي، ينبغي أن يكون توزيع المتوســطات لمتوســط الطــول توزيعــاً طبيعيًا (خاصةً إذا قمنا بســحب أو أخذ عدد كاف مــن العينات).

(٢) بوتستراب (Bootstrap) هي تقنية لإعادة المعاينة والتي هي حالة خاصة من أساليب المحاكاة بطريقة مونت كارلو وذلــك للحصــول على مقدرات أكثر دقة عندما يكون حجم العينة صغــيراً، أو تعذر تحديد التوزيع الاحتمالي للمقدر محل الاهتمام. والفكرة الأساســية لهــذه التقنية تتضمن القيام بتوليد عدد كبير من العينات المســحوبة مع الإرجاع من بيانات العينة الأصلية. كذلك، فهذه الطريقة تزودنا بالخطأ المعياري للمقدرات بدون أن يتطلب ذلك حســابات نظرية، وهذا يوفر الوقت والجهد (المترجمَين).

وكما هو الحال مع أي توزيعات طبيعية، يمكننا تحديد الحدود التي كانت بالنسـبة لـ ٩٥٪ المركزية من متوسـط الأطوال، مـع تحديد قيم العينات المقابلة للحـدود الدنيا والعليا لهذه المنطقة. لذلك؛ إذا كان المتوسـط الكلي (على سـبيل المثال، متوسـط كل متوسطات العينات الفرديـة) هو ٦٨ بوصة، فيمكننـا إيجاد الحدود العليا والدنيا لفـترة الثقة البالغة ٩٥٪ حول تقدير هذه النقطة. وتتبع بوتسترابنج Bootstrapping نفس المنطق مع اختلاف مهم للغاية. فبدلاً مـن أخذ عينات متعددة من المجتمع، فإنه يحـاكي هذه العملية من خلال أخذ عينات فرعيـة متعـددة من عينة واحدة متوفرة. وفي الأسـاس، يتم التعامل مـع العينة التي تكون في متنـاول اليد كمجتمع ويتم السـحب مع الإرجـاع replacement لعدد كبير من "عينات" مختلفة منها. بهذه الطريقة، يتم إنشـاء إصدارات أو نسخ مختلفة من العينة ويمكن حساب معامـل ألفا في كل منها، مما يؤدي إلى توزيع درجـات معامل ألفا التي يمكن أن تكون بمثابة أساس لتحديد فترات الثقة.

وتقدم باديلا وآخرون (Padilla et al., 2013) وصفاً مفصلاً لطريقة بوتسـترابنج لحساب فترات الثقة لمعامل ألفا باسـتخدام بيانات المحاكاة. وكوسيلة لفهم فائدة طريقة بوتسترابنج، يمكن أن تكون بيانات المحاكاة مفيدة بشـكل خاص لأن مجموعة البيانات يمكن إنشاؤها مع الثبات «الحقيقي» للمقياس الافتراضي المحدد مسـبقاً من قبل الباحثين. ويمكن تحديد قيمة درجة حقيقية تعسفية كما يمكن توليد بيانات «ملاحظة» تختلف عن الدرجة الحقيقية بطرق تخضع تماماً لسيطرة أو ضبط الباحثين. وبالتالي، يمكن للباحثين التحكم في كل من مدى ثبات المقيـاس الافتراضي من البدايـة وخصائص التوزيع (مثل الالتـواء والتفلطح) لمجموعة بيانات المحـاكاة. وبعد ذلك، يمكـن للباحثين تنفيذ طرق مختلفة للبوتسـترابنج وتحديد أيهم يوفر تمثيلاً أكثر دقة للقيم «الحقيقية» التي قاموا ببنائها من بيانات المحاكاة. واستخدمت «باديلا وآخـرون» (٢٠١٣) هـذا المدخل لمقارنة طرق مختلفة لتقدير فـترات الثقة لمعامل ألفا تحت ظروف بيانات مختلفة (على سبيل المثال، بيانات ذات توزيع غير طبيعي). وتشير نتائجها إلى أن بعض الطرق تعمل بشكل جيد عبر مجموعه متنوعة من الظروف، في حين أن بعض الطرق الأخـرى تعتمـد بدرجه أكبر على شروط توزيعية معينه للتأكد من دقتها. وبشـكل أعم، فإن النتائج التي توصلوا إليها، اسـتناداً إلى عمليات المحاكاة، توفر رؤى تفصيليه عن الوقت الذي يمكن فيه تطبيق مدخل البوتسترابنج لإيجاد فترات الثقة لمعامل ألفا كتقدير لمدى الثبات.

تتمثل أحد الجوانب السلبية المحتملة لاستخدام المحاكاة في أنه من السهل وغالبًا ما يكون من المغري محاكاة خصائص التوزيع المتطرفة التي تنتج اختلافات كبيرة بين المعلمات المقدرة والمعلمات النظرية مثل الثبات. على الرغم من أن أهمية أسلوب المحاكاة، إلا أن أكثر ما يهتم به الباحثون هو خصائص بياناتهم الفعلية. لحسن الحظ، يمكن أيضًا استخدام مدخل البوتستراينج bootstrapping مع البيانات الفعلية لتحديد فترات الثقة حول قيمة معامل ألفا التي تم الحصول عليها. لم يتم استخدام Bootstrapping على نطاق واسع بهذه الطريقة في الماضي، ويرجع ذلك جزئيًا إلى أن الأدوات الحسابية لإنشاء نماذج فرعية من bootstrap ولحساب فترات الثقة لم تكن متوفرة بسهولة. ولقد تغير هذا إلى حد ما مع ظهور برنامج التحليل الإحصائي R (Venables, Smith, & the R-Core team, 2015) وهي بيئة برمجية قوية ومتعددة الاستخدامات تمكن المستخدمين من أداء مجموعة واسعة من عمليات التحليل والرسومات البيانية. وبرنامج التحليل الإحصائي R مجاني ويمكن الوصول إليه من خلال الرابط التالي :http://www.r-project.org. هذا الموقع يوفر أيضاً الوثائق. وداخل بيئة البرنامج الإحصائي R، يوجد عدد من الحزم المتاحة والتي هي في الأساس مجموعات من البرامج الفرعية المصممة حول تحقيق أهداف تحليلية أو رسومية محددة. وهذه أيضًا متاحة مجانًا ويمكن الوصول إليها من مواقع نسخ متطابقة متعددة من داخل بيئة البرنامج الإحصائي R.

معامل أوميغا (ω):

على الرغم من أنه من الواضح إمكانية تعزيز فائدة معامل ألفا باستخدام الارتباطات الرباعية /المتعددة الأقسام أو عن طريق تحديد فترات الثقة، يوصي العديد من المتخصصين في منهجية البحوث بمدخل بديل يعالج محددات أو قيود معامل ألفا بشكل كامل. هذا العلاج أو الوسيلة تتمثل في معامل أوميغا (ω)، وهو تقدير بديل للثبات. ومنطق معامل أوميغا مشابه لمنطق معامل ألفا. وكلاهما يعرفان الثبات أساساً على أنه نسبة تباين الدرجة الحقيقية، مقارنة بالتباين الكلي الملاحظ. ومع ذلك، يستخدم معامل أوميغا طريقه مختلفة لحساب التباين ذي الصلة. في حين يشتق معامل ألفا تقديرات التباين من مصفوفة التغاير (أو الارتباط) للبنود التي تشكل مقياساً، فإن معامل أوميغا يستخدم مصفوفة من تشعبات

البنود على عامل مشــترك وحيد تشترك فيه البنود. وستتم مناقشة التحليل العاملي بشكل أكثر شــمولاً في فصل لاحــق، ولكن في الوقت الراهن، يمكننا أن نؤكد أن أحد اســتخداماته هــو تقدير الارتباط بــين كل مجموعه من البنود، من ناحية، وعامل يمثل المتغير الأســاسي المشــترك بينهما، مــن ناحية أخرى. وهذه الارتباطات هي تشــبعات البنــود على العامل. ويســتند معامل أوميغا على افتراضات النموذج المتقــارب أو التقاربي congeneric (الذي تمت مناقشته بإيجاز في الفصل الثاني) بدلاً من افتراضات نموذج «تاو» المتماثل أو المترادف في الأســاس التي يعتمــد عليها معامل ألفــا. والاختلاف الرئيسي بين الافتراضات بالنســبة للنموذجين هو أن النموذج التقاربي لا يتطلب أن تكون التغايرات للبنود مع عاملها المشترك (الوكيل لتحقيق الدرجة الحقيقية) متســاوية أساســاً. وكل ما هو مطلوب فعلاً هو أن كل البنود تعكس متغيراً أساسياً مشتركاً واحداً يؤدي إلى عامل مشترك واحد فقط.

ومن الناحية النظرية، فإن اســتخدام معامل أوميغا يعد أمراً جذاباً للغاية. فهو يعتمد على نفس منطق معامل ألفا، أي تقســيم التباين الملاحظ إلى جزء يعكس الدرجة الحقيقية والباقي يعكس الخطأ. ومن ثم، فإن النسبة الكلية (أي تباين الدرجة الحقيقية زائد تباين الخطــأ) التي تعــزى إلى الدرجة الحقيقية ســتقدر الثبات. ولكنه يفعــل ذلك دون قيود ومحددات افتراض لنموذج «تاو» المتماثل أو المترادف في الأســاس. ولا يزال افتراض أحادية البعــد (أي إن جميع البنود تعكس متغيرًا ســببيًا واحدًا مشــتركًا) مطلبًا، ولكن تم إهمال والتخــلي عن القيد الإضــافي والمتمثل في أن تكون التغايرات بين البنــود والدرجة الحقيقية (والتي تقارب تحميلات أو تشعبات البنود على العامل) متساوية بشكل أساسي.

وعلى الرغم من هذه المزايا النظرية، فإنه لم يتم اعتماد معامل أوميغا على نطاق واسع بســبب التعقيد الكبير في طريقة حســابه. عــلى عكس معامل ألفا، والذي يمكن حســابه باستخدام قلم رصاص وورق بمجرد حساب بعض الإحصائيات الأساسية (مثل تباينات البنود والتباين الكلي للمقياس)، أما اســتخدام معامل أوميغا فإنه ينطوي على زيادة عدد الأرقام بشكل كبير. ولم تكن الإجراءات الروتينية الملائمة لحساب معامل أوميغا متوفرة على نطاق واسع في حزم برامج إحصائية شائعة، ولكن هذا يتغير الآن.

مرة أخرى، يقدم البرنامج الإحصائي R حلاً مجانياً ويمكن الحصول عليها بسهوله للمعضلة الحســابية. على الرغم من أن إدراج وصفٍ مفصلٍ للإجراء التحليلي في البرنامج الإحصائي R

هــو خارج نطاق هذا الكتاب، فــإن "بيترز" (Peters, 2014) يصــف كيفية إعداد البرنامج الإحصــائي R بحيث تكون هناك حاجة فقط لكتابة ســطرين من التعليمات البرمجية code لحساب كل من معامل ألفا ومعامل أوميغا، كاملة مع حساب فترات للثقة لكل منها. وهذا يســمح للباحث ليس فقط الحصول عــلى معامل أوميغــا ولكن أن يقارنه مع معامل ألفا. وبالنســبة لأولئك الذين يفضلون نظرة أكثر تفصيلاً إلى حد ما لعمليات البرنامج الإحصائي R والذين قد يحتاجون إلى بعض التوجيه فيما يتعلق باستيراد ملف بيانات من مصادر مثل إكسل أو برنامــج SPSS إلى البرنامج الإحصائي R، فقد وصف دان وآخرون (Dunn et al., 2014) وصفاً جزئياً عملية آلية مبسطة لحوسبة معامل أوميغا مع فترات الثقة. ومن الجدير بالذكر أن مدخل البوتســترابنج المســتخدمة في هذه الحسابات يتضمن إنشــاء ما يصل إلى ١٠٠٠ عينة فرعيه من مجموعه البيانات الأصلية وأداء الحســابات على كل عينة من هذه العينات الفرعية. وهكذا، فإن العملية برمتها ســوف تستغرق وقتاً أطول بكثير حتى تكتمل وتنتهي من الحساب البسيط لمعامل ألفا باستخدام الحزم الإحصائية القياسية. والوقت الذي ينطوي عليه الأمر هو دليل على حقيقة أن هذه الإجراءات أصبحت متاحه على نطاق واســع فقط مع ظهور أجهزة الحاسب الآلي وبيئات البرمجيات المتقدمة مثل البرنامج الإحصائي R.

وفي رأيــي، فإن معامل ألفــا يحتفظ بالفائدة والمنفعة. وهو كتقدير للحد الأدنى للثبات، فإنه يعد تقديراً متحفظاً للثبات. وعلاوة على ذلك، عندما يتم توخي الحذر في وضع واختيار البنــود وأن تكــون متوافقة مع افتراض أحادية البعد ونمــوذج «تاو» المتماثل أو المترادف في الأساس، فإن العديد من المخاطر المحتملة لاستخدام معامل ألفا كمؤشر على الثبات تتضاءل أو تقل. أخيراً، تم استخدام معامل ألفا كرونباخ على نطاق واسع، واستندت معايير الباحثين حــول متى تكون أدوات القياس ثابته بدرجة كافية إلى حد كبير على اســتخدام معامل ألفا. وإذا كان معامل ألفا متحفظاً بعض الشيء في بعض الحالات، ثم بالنســبة لأي مستوى معين من الثبات، فإن بعض المؤشرات البديلة قد تنتج عنها قيمة رقمية أعلى من قيمة معامل ألفا. قد يكون هذا مربكاً. أي مقياس معين تحت شروط وظروف معينة سيكون له ثبات محدد، ولا تغير طريقة تقدير ذلك الثبات التي تســفر عن قيمة رقمية أكبر من ثبات المقياس. إذا حقق مقياس معين في مناســبة معينة قيمة لمعامل ألفا تســاوي ٠,٨٠ وقيمة لمعامل أوميغا بقيمــة ٠,٨٢، فإن ثبات المقياس لم يتغير نتيجة للحســاب الأخير باســتخدام معامل أوميغا.

صحيح أن قيمة معامل أوميغا قد تكون أكثر دقة من الناحية الرقمية، لكن هذا لا يعني أن ثبات المقياس قد تغير، بل إن المقياس المستخدم لقياس هذا الثبات قد تغير. وبعبارة أخرى، فـإن القيمة العددية الأعلى البالغة ٠,٨٢ لمعامل أوميغا، في هــذه الحالة الافتراضية، تعادل أساسًا (أي أنها تمثل نفس درجة الثبات) القيمة ٠,٨٠ لمعامل ألفا. ومن ثم فإن اعتماد وتبني معيار جديد لتمثيل الثبات يتطلب إعادة معايرة معاييرنا الحالية القائمة على أســاس معامل ألفا، وهذا لن يكون تحولًا خطيًا بســيطًا؛ لا يمكن للمرء أن يحول بين تقدير الحد الأدنى glb أو قيمة معامل أوميغا ومعامل ألفا ببســاطة بمجرد الإضافة أو الضرب في مقدار ثابت. (في الفصل التالي، أناقش حججًا مماثلة ضد تصحيح الارتباطات الضعيفة عند دراســة الصدق). ولأن معامل أوميغا سيكون أكثر دقة (على سبيل المثال، إذا كان التفكير في معامل الثبات من حيث نســبة التباين في المقياس المشترك مع الدرجة الحقيقية)، فسيكون من المفيد في نهاية المطاف التعود على المقياس الجديد. ومع استمرار العمل على بدائل مختلفة لمعامل ألفا، قد نصل قريبًا إلى وقت يتم فيه تبرير مزايا هذه البدائل باعتمادها على نطاق أوسع. ومع ذلك، من المهم في غضون ذلك إدراك أن القيم العددية الأعلى لمعامل أوميغا بالنســبة إلى معامل ألفا لنفس البيانات الناتجة عن نفس المقياس هي إعادة معايرة وليست زيادة في الثبات.

الثبات القائم على الارتباطات بين درجات المقياس:

هناك بدائل لمعامل ألفا كمؤشر على الثبات. قد تتضمن هذه الأنواع من حساب الثبات وجود نفس مجموعة الأشــخاص الذين يكملون نســختين منفصلتين مــن مقياس أو نفس النسخة من المقياس في مناسبات متعددة.

ثبات النماذج البديلة:

إذا كان هناك شكلان متماثلان بشكل تام للمقياس، يمكن حينئذ حساب الارتباط بينهما طالما أن نفس الأشــخاص يكملون الاســتجابة على كلا الشكلين المتماثلين. على سبيل المثال، افترض أن الباحث قام أولاً بتطوير مجموعتين متماثلتين من البنود التي تقيس رغبة المرضى في التحكم عند التفاعل مع الأطباء، ثم قام بإدارة وتطبيق مجموعتي البنود لمجموعة واحدة مــن المرضى، ثم، أخيرًا، قام بحســاب الارتباط بين درجات مجموعــة واحدة من البنود مع

درجات مجموعة البنود الأخرى. فإن هذا الارتباط سيكون هو ثبات النماذج البديلة. تذكر أن النــماذج المتكافئة أو المتوازية تتكون من بنود، كلها (إما داخل النماذج أو بينها) تؤدي عملاً جيداً بنفس القدر لقياس المتغير الكامن. وهذا يعني أن كلا الشكلين من المقياس لهما معاملات ألفا، ومتوســطات حســابية، وتباينات متماثلة، ويقيسان نفس الظاهرة. وتتكون النــماذج المتكافئة، في جوهرها، من مجموعة واحدة من البنود التي تم تقســيمها بطريقة عشــوائية إلى مجموعتين فرعيتين تشكلان النموذجين المتماثلين والبديلين للمقياس. في ظل هذه الظروف، يكون الارتباط بين أحد النماذج والنموذج الآخر مكافئًا أو مســاوياً لارتباط أي من النموذجين مع نفسه، حيث إن كل نموذج هو بديل مكافئ للآخر.

ثبات التجزئة النصفية:

تتمثل مشكلة النماذج البديلة للثبات في أننا عادةً لا نملك نسختين من المقياس تتوافقان تمامًــا مــع افتراضات الاختبارات المتكافئــة أو المتوازية. ومع ذلك، هنــاك تقديرات أخرى للثبــات تطبق نفس النوع من المنطق على مجموعة واحدة من البنود. ونظرًا لأن النماذج البديلة تتكون أساسًــا من مجموعة واحدة من البنود التي تم تقســيمها إلى قســمين، فإنه يترتب على ذلك أنه يمكننا (أ) أخذ مجموعة البنود التي تشــكل مقياسًا واحدًا (أي مقياس لا يحتوي على أي نموذج بديل)، و(ب) تقسيم مجموعة البنود هذه إلى مجموعتين فرعيتين، و(ج) إيجاد الارتباط بين المجموعات الفرعية لتقييم الثبات.

ويُسمى مقياس الثبات من هذا النوع باسم ثبات التجزئة النصفية **Split-half reliability**. ويُعد ثبات التجزئة النصفية في الواقع طبقياً class وليســت نوعاً واحداً من الطرق الحســابية نظراً لوجود مجموعة متنوعة من الطرق التي يمكن بها تقســيم المقياس إلى نصفين. وتتمثل إحــدى الطــرق في مقارنة النصف الأول من البنــود بالنصف الثاني. قد يمثــل هذا النوع من تقســيم المقياس إلى نصفين هما نصف أول، ونصف أخير إشكالية، لأن عوامل أخرى غير قيمة المتغير الكامن (بمعنى آخر، مصادر الخطأ) قد تؤثر على كل مجموعة فرعية بشــكل مختلف. على ســبيل المثال، إذا كانت البنود المكونة للمقياس المحدد منتشرة ومتناثرة خلال اســتبيان مطول، فقد يكون المجيبون أكثر تعبًا أو إرهاقاً عند إكمال النصف الثاني من المقياس. وعندئذٍ، سيختلف الإرهاق أو التعب بشكل منتظم بين النصفين، مما يجعلهما أقل تشابهاً على ما يبدو.

ومع ذلك، فإن عدم التماثل أو الاختلاف لن يكون سـمة من سـمات البنود في حد ذاتها بقدر مـا هي موضعها في ترتيـب البند من المقياس. ومن العوامل الأخـرى التي قد تحدث الفرق او الاختـلاف بين البنـود في النصف الأول والبنود في النصف الأخير، تأثير الممارسـة (حيث قد تتحسـن إجابة المسـتجيبين على البنود أثناء سـيرهم في عملية الإجابة)، أو الإخفاق والفشل في إكـمال مجموعه البنـود بأكملها، أو ربما حتى شيء عادي مثـل التغييرات في جودة طباعة الاسـتبيان من الأمام إلى الخلف. وكما هو الحال بالنسـبة لحالة التعب أو الإجهاد، فإن هذه العوامل سـتخفض الارتباط بين النصفين بسـبب الترتيب الذي عُرضت به بنود المقياس وليس بسبب نوعيه أو جودة بنود المقياس. ونتيجة لعوامل مثل هذه، فإن قياس قوة العلاقات بين البنود قد يكون معقداً بسـبب عدم اتصال الظروف مباشرةً بجودة البند، مما يؤدي إلى تقييم الثبات بصوره خاطئة.

ولتجنـب بعض المخاطر المرتبطة بترتيب البند، يمكن للمرء أن يقيم نوعاً آخر من ثبات التجزئة النصفية المعروفة بثبات البنود الفردية والزوجية odd-even reliability. في هذه الحالـة، تتم مقارنه المجموعة الفرعية مـن البنود ذات الترقيم الفردي بالبنود ذات الترقيم الزوجي. وهذا يضمن أن كلاً من المجموعتين الفرعيتين من البنود تتكون من عدد متسـاو من كل قسم (أي بداية ومنتصف ونهاية) من المقياس الأصلي. على افتراض أن ترتيب البند غير ذي صلة (على عكس الترتيب «السهل إلى الصعب» الشائعة في اختبارات التحصيل، على سـبيل المثال)، فإن هذا الأسـلوب يتجنب العديد من المشاكل المرتبطة بتقسيمات النصف الأول مقابل النصف الثاني.

ومـن الناحية النظرية، هنـاك العديد من الطرق الأخرى للوصـول إلى الثبات بطريقة التجزئة النصفية. هناك بديلان للطرق التي تمت مناقشتها أعلاه لتكوين مجموعات البنود الفرعية هما النصفان المتوازنان balanced halves والنصفان العشوائيان random halves. في الحالة الأولى، يمكن للمرء أن يحدد بعض خصائص أو سمات البند التي يحتمل أن تكون مهمـة (مثـل، الصياغة في حد ذاتها، أو طول البند، أو ما إذا كان نوع معين من الاسـتجابة يشـير إلى وجود أو غياب السـمة المعنية). ثم يتم تشـكيل نصفي المقياس بحيث تكون الخصائص ممثلة بالتسـاوي في كل نصف. وهكذا، يمكن للباحث أن يقسـم البنود بحيث يكـون لكل مجموعة فرعيـة نفس عدد البنود التي تمت صياغتها في القسـم أو المجموعة

الأولى ونفـس عدد العنـاصر القصيرة وما إلى ذلك. غير أنه عنـد النظر في خصائص البنود المتعددة، قد يكون من المسـتحيل تحقيق التوازن بين نسبة بند دون أن يكون التوازن بين بند وبند آخر أمراً مسـتحيلاً. وهذا هو الحال، على سـبيل المثال، إذا كانت هناك بنود في القسـم أو النصف الأول أكثر من البنود القصيرة. ومن شـأن إيجاد تـوازن لهذه الخاصية الأخيرة أن يستلزم اختلال التوازن في الخاصية الأولى أو السابقة. أيضًا، قد يكون من الصعب تحديد خصائص البنود التي ينبغي موازنتها.

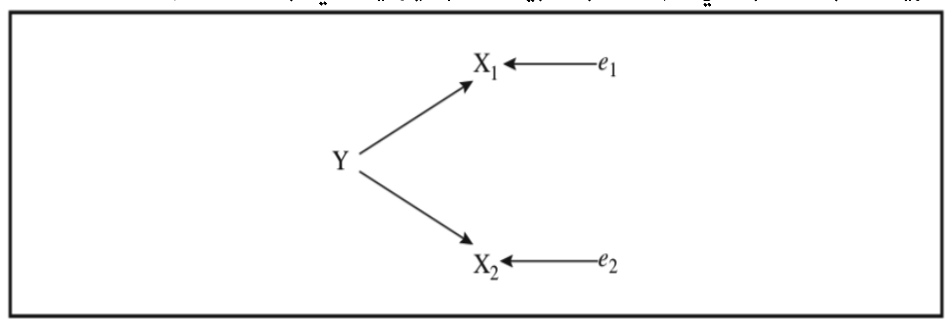
يمكن أن يحصل الباحث على نصفين عشـوائيين بمجرد التوزيع العشـوائي لكل بند من البنود على إحدى المجموعتين الفرعيتين اللتين سيتم في النهاية حساب الارتباط مع بعضهما البعض لحسـاب تقدير الثبـات. ويتوقف مدى نجاح هذا العمل عـلى عدد البنود وعدد الخصائص موضع الاهتمام، ودرجة الاسـتقلال بين الخصائص. على أمل أن يسفر عدد قليل من البنود، التي تختلف وفقًا لعدة أبعاد مترابطة، عن مجموعات قابلة للمقارنة من خلال التوزيع العشوائي، وهذا أمر غير واقعي. ومن ناحية أخرى، فإن التوزيع العشوائي لمجموعة مكونة من ٥٠ بنداً تتفاوت وتختلف فيما يتعلق بسـمتين أو ثلاث سـمات غير مترابطة إلى فئتين قد يؤدي إلى مجموعات فرعية قابلة للمقارنة فيما بينها بشكل معقول.

وتعتمـد أفضـل طريقة يتم بها الحصـول على نصفين مجزأين عـلى الموقف أو الحالة الخاصـة بها. وأهم مـا في الأمر هو أن يفكر الباحث في طريقة تقسـيم البنود التي يمكن أن تـؤدي إلى مجموعات فرعية غير متكافئة والخطـوات التي يمكن اتخاذها لتجنب ذلك. إن التبريـر المنطقي وراء كل من ثبات التجزئـة النصفية وثبات النماذج البديلة هو امتداد طبيعي لنموذج الاختبارات المتكافئة (المتوازية).

وعلى الرغم من أننا عندما ناقشنا هذا النموذج في البداية، اعتبرنا أن كل بند هو اختبار، يمكن للمرء أن ينظر أيضاً إلى مقياس (أو نصفين من مقياس) يتوافق مع النموذج كاختبار. ولذلـك، يمكننا تطبيق المنطق الذي اسـتخدمناه في حاله البنود المتعـددة على حالة اثنين مـن النماذج البديلة أو نصفين مـن المقياس. فكر في اختبارين (نصفي المقياس أو النماذج البديلة) تحت افتراضات الاختبارات المتكافئة، كما هو موضح في الشكل رقم (٣-٣).

والسـبيل الوحيد الذي يربط بين المكونين الاثنين للمقياس هو المسـارات السـببية من المتغـير الكامن إلى كل نصف مـن نصفي المقياس. وهكذا، فإن ناتج قيم هذه المسـارات

يساوي الارتباط بين الاختبارات. وإذا كانت قيم المسار لابد أن تكون متساوية (وهي تكون كذلــك تحــت افتراضات هذا النموذج)، فــإن الارتباط بين الاختبارات يســاوي مربع قيمة المسار من المتغير الكامن إلى أي من الاختبارين. مربع ذلك المسار (على افتراض أنه معامل مســار معياري) هو أيضاً نسبة التباين في أي اختبار يتأثر بالمتغير الكامن. وهذا بدوره هو تعريف الثبات. وبالتالي، فإن الارتباط بين الاختبارين يساوي ثبات كل منهما.



شكل (٣-٣): مخطط مسار يوضح علاقات نصفي المقياس منفصلين (X1) و(X2) بالمتغير الكامن المشترك بينهما

وفي حين أن الاختبارات المشــار إليها في الفقرة الســابقة هي عبارة عن نسختين كاملتين من المقياس في حالة النماذج البديلة، إلا أنهما عبارة عن نصفين للمقاييس في حالة التجزئة النصفيــة. وبالتالي، فإن الارتباط بين نصفي المقياس يؤدي إلى تقدير الثبات لكل نصف من المجموعة الكاملة من البنود، وهو ما يقلل من تقدير ثبات مجموعة البنود بأكملها. ويمكن حســاب تقدير ثبات المقياس بأكمله، استنادًا إلى ثبات جزء من المقياس، باستخدام معادلة ســبيرمان براون Spearman-Brown formula، التي تمت مناقشتها سابقًا في هذا الفصل. تذكر أنه وفقًا لهذه المعادلة، فإن معامل الثبات ألفا يساوي:

$$\alpha = \frac{k\bar{r}}{1 + (k-1)\bar{r}}$$

حيث k هي عدد البنود المعنية وهو متوســط ارتبــاط بين بند ما مع أي بند آخر (أي، متوســط الارتباط بين البنود). إذا كنت قد حددت مدى ثبات مجموعة فرعية من البنود (على سبيل المثال، من خلال طريقة التجزئة النصفية) وتعرف عدد البنود التي يستند إليها الثبات (على ســبيل المثال، نصف العــدد في المقياس بأكمله)، فيمكنك اســتخدام المعادلة

لحســاب متوســط ارتباط بــين بند ما مع أي بنــد آخر. بعد ذلك، يمكنك بمســاعدة قيمة هــذه وعــدد العناصر في المقياس بأكملهــا لتعويض في المعادلة. وســتكون النتيجة تقديراً لثبات المقياس بالكامل، اســتناداً إلى قيمة الثبات المحســوبة لنصفي المقياس المجزأ. وهذا يعمل على تبسيط الأمور إذا قمت بإجراء بعض العمليات الجبرية (الحسابية) على معادلة سبيرمان - براون لوضعها في الشكل التالي:

$$r = \frac{r_{yy}}{\left[k-(k-1)r_{yy}\right]}$$

حيث ryy هو ثبات البند المعني. على سبيل المثال، إذا علمت أن ثبات التجزئة النصفية لنصفي مقياس مكونين من تسعة بنود تساوي ٠,٩٠، يمكنك حساب $\overline{r}$ على النحو التالي:

$$\overline{r} = \frac{.9}{[9-(8)(.9)]} = .5$$

ويمكنك بعد ذلك إعادة حســاب الثبات للمقياس الكامل المكون من ١٨ بنداً باستخدام تســاوي (٠,٥)، وعدد البنود k يســاوي (١٨) في معادلة سبيرمان - براون. وعلى ذلك، فإن تقدير الثبات للمقياس الكلي يكون:

$$\frac{18\times.5}{1+(17\times.5)}$$

الذي يساوي ٩ / ٩,٥، أو٠,٩٤٧. لاحظ أن زيادة عدد البنود قد زاد من الثبات. إن إلقاء نظرة سريعة على معادلة سبيرمان - براون من شأنه أن يجعل من الواضح أنه، إذا كان كل شيء متساوياً، فإن المقياس الأطول سيكون دائمًا أكثر ثباتاً من المقياس الأقصر. قد لا يكون هذا واضحًا بشــكل حدسي أو بديهي. إذا كان لكل بند بعض الإشــارات وبعض الضوضاء، ألا تؤدي إضافة البنود إلى زيادة الضوضاء وكذلك الإشــارة؟ الجواب هو، نعم، ســوف تزيد البنــود ذلك؛ ومع ذلك، يبــيّن لورد ونوفيــك (Lord and Novick, 2008) أنه بينما يزداد الخطــأ خطيّــا، تزداد الإشــارة signal أضعافاً مضاعفة عند إضافة البنــود (ص. ٨٦) – أي إن تباين الدرجة الحقيقية يتراكم بشــكل أسرع من تباين درجة الخطأ مع إطالة الاختبار.

وهذا مماثل للطريقة التي يصبح بها أي تقدير لمعلمة المجتمع (على سبيل المثال، متوسط وزن المراهقــين في الولايــات المتحدة) تدريجيــاً أكثر دقة (أي أقــرب إلى الدرجة الحقيقية للمجتمع) مع زيادة عدد الأفراد الذين تم أخذهم في العينة.

اتفاق المُقيّمين inter-rater agreement:

تضمنت أســاليب تقييم الثبات التي ناقشــناها حتى الآن على البنود كمؤشرات. ومن المؤكد أن هذا مناســب لكتاب عن وضع وتطوير المقياس؛ غير أنه، في العديد من السياقات البحثيــة، يعمــل المقيمــون أو المحكمــون كمؤشرات. عــلى الرغم من عــدم وجود صلة وثيقــة لذلك بتطوير المقياس، إلا أن مناقشــة موجزة لاتفاق المقيمــين (انظر دي فيلييس DeVellis, 2005، للحصــول على نظــرة عامة حول هذا الموضوع) تكون مفيدة لأنها يمكن أن تؤكد التشــابه المفاهيمي الذي تشترك فيه مع تقييم الثبات القائم على البند. ففي كلتا الحالتين، يفرض المنطق الأساسي أن المؤشرات التي تتقاسم قضية مشتركة يجب أن ترتبط مع بعضهــا البعض. في حالة المقيمين أو المحكمين، إذا كانت الدرجات الناتجة عن ملاحظاتهم تعكــس خصائص المثــير أو المنبه الذي يتم ملاحظته بدلاً مــن خصائص المحكمين، فينبغي عندئذ أن تتفق تلك الدرجات. ويمكن أن تختلف طبيعة هذا الاتفاق تبعاً لأهداف الباحث والمنهج المتبع في تقدير اتفاق المقيمين.

وفي بعض الحالات، ســيحتاج الباحث إلى اتفاق صارم من أجل النظر في التقديرات التي يمكــن التعويــل عليها والوثوق بها. وكمثال على ذلك، فإن المدى الذي يمكن فيه لاثنين من حــكام البيســبول (كرة القدم الأمريكية) أن يحددا بشــكل ثابت متــى يكون ضرب الكرة صحيحاً أو خطأً يتوقف على مدى اتفاقهما على سلســله من الأحداث أو المناسبات. يجب أن يكون أي ضرب للكرة إما صحيحاً أو خطأً وفقاً لقواعد البيسبول. ومن ثم، فإن الاتفاق المطلق يكون كافياً كوسيلة لقياس مدى ثبات تقديرات الحكام. وفي ظروف أخرى، قد يقوم المقيمون بتقييم ظاهره تختلف على طول سلســلة متصلة. على سبيل المثال، من المحتمل ان يســتخدم الحكام في ســوق ولاية ما يقومون بتقييم نوعية أو جــودة الفطائر المخبوزة محلياً مقياســاً له العديد من الاســتجابات بدلاً من التقسيم الثنائي البسيط (جيد أو سيئ). وفي مثــل هذه الحالــة، قد لا يكون من الضروري التوصل إلى اتفــاق دقيق حول الدرجة

لتحديد أن تقديرات الحكام المتعددين تعكس في الواقع سمة الفطائر وليست مجرد سمات الحكام أنفسهم. وعلى الرغم من أن أحد الحكام قد يكون أكثر تساهلاً من قاض آخر غيره وبالتالي يعطي تقييمات أعلى باستمرار، إلا أن الدرجات بين الحكام ينبغي أن تتماثل تقريباً إذا كان يجب اعتبارها تتمتع بالثبات. والواقع، قد يكون هذا الأســاس الأفضل للحكم على الثبــات في هذه الظروف بدلاً من الإصرار على الاتفاق التــام. والتقييمات التي تم تعيينها حتى يكون ترتيب الفطائر المحرز أو المســجل من الأفضل إلى الأســوأ متشــابهاً بين الحكام توحي بأن سمات الفطائر وليست سمات الحكام هي التي تحدد الدرجات التي تم تعيينها. ومع ذلك، إذا كانت معايير تســجيل الحكام تمت معايرتها بشكل مختلف إلى حد ما (مثلاً، إذا اعتــبر أحد الحكام أن الدرجة خمســة تمثل متوســط جودة الفطــيرة، بينما أعطى آخر للفطيرة متوســطة الجودة ست درجات)، فقد لا يتفق الحكام أبدًا في الواقع. ومن ثم، قد لا يكون الاتفاق التام معيارًا مناسبًا لتقييم الثبات فيما بين الحكام في مثل هذه الحالة.

وفي ورقــه مؤثــره للغايــة، وصــف باتريــك ا. شروت وجوزيــف ل. فلييــس (Patrick E. Shrout and Joseph L. Fleiss, 1979) الطــرق المختلفــة لتقييــم اتفــاق المقيمين في ظل ظروف مختلفة. وأكدوا على أن السمات المهمة لنوايا الباحث تحدد أفضل طريقه لتقييم الثبات فيما بين المقيمين. ولأغراض المناقشة الحالية، فإن أهم سمة ناقشوها هي ما إذا كان الاتفاق الدقيق أو مجرد التطابق بين المحكمين يمثل أفضل دليل على اتساق المقيمين. وفي مثال الحكام، من الواضح، أن الاتفاق المطلق ســيكون ضرورياً لاســتنتاج أي من الحكام يتمتع بالثبات. إن نسخة معامل الارتباط داخل الطبقة ICC المناسبة في هذه الحالة تعادل أو تساوي معامل كوهن كابا Cohen's kappa (k) (كوهين؛ Cohen, 1960). ويحــدد هــذا المدخل إلى أي مدي يتجاوز تكرار الاتفاقــات الدقيقة بين المحكمين ما يمكن توقعــه عن طريــق الصدفة. وعلى النقيــض من الحكام، يمكن لقضــاه الحكم على جودة الفطــيرة أن يبرهنــوا على أنهم كانوا يقومــون بإصدار أحكام تتمتــع بالثبات دون الاتفاق الفعــلي على ذلك بدقة. نموذج معامل الارتباط داخل الطبقة ICC الذي يعبر عن الاتفاق في ظل هذه الظروف يعادل معامل ألفاكرونباخ.

وعلى ذلك، يمكن اســتخدام معامــل ألفاكرونباخ عندما يهتم المرء بالمدى الذي يســفر فيــه أي من البنود أو المقيمين عن درجات تكون متوافقــة أو متطابقة. وفي كلتا الحالتين،

ينشأ التوافق أو التطابق (أي الارتباطات بين البنود وبعضها البعض أو بين المقيمين) بسبب التأثير المشـترك للمتغير الأساسي على المؤشرات، سواء كانت هذه الأخيرة عبارة عن بنود أو مقيمين. وتشـكل نسـبة تباين الدرجة الملاحظة التي يمكن إرجاعها إلى الدرجة الحقيقية للمتغير الكامن ثبات المؤشرات في كلتا الحالتين. وتؤكد هذه المقارنة بين أنواع مختلفة من المؤشرات (المقيمين مقابل البنود) تفسير الفائدة العامة لمعامل ألفا كمؤشر على الثبات.

الاستقرار الزمني:

هنـاك طريقـة أخـرى لحسـاب الثبـات تنطـوي عـلى الاسـتقرار الزمنـي للقيـاس temporal stability، أو مدى بقاء الدرجات أو النتائج ثابتة (مسـتقرة) من مناسـبة أو من حالة إلى أخرى. ويُعد ثبات إعادة الاختبار Test-retest reliability هي الطريقة المستخدمة عـادةً لتقييم ذلك النوع من الثبات. وفي بعض النواحـي، يكون ثبات إعادة الاختبار مماثلاً ومشابهاً لثبات التجزئة النصفية وثبات النماذج البديلة، والتي تمت مناقشتها سابقاً. وبالعودة إلى مثال سابق، افترض أنه بدلاً من تطوير مجموعتين من البنود لقياس رغبة المرضى للتحكم أو السيطرة عند التفاعل مع الأطباء، وضع باحثنا الافتراضي مجموعة واحدة من البنود. ويمكن تطبيق هذه البنود على مجموعة واحدة من المرضى في مناسـبتين منفصلتين، ويمكن حسـاب ارتبـاط الدرجات التي حصلوا عليها في المناسـبة الأولى بتلـك الدرجات التي حصلوا عليها في المناسبة الثانية. والأساس المنطقي الذي تستند إليه القرارات المتعلقة بالثبات من هذا النوع هـو أنه إذا كان المقياس يعكـس حقًا (في الواقع) مغزى لبنـاء مفاهيمي محدد، فإنه يجب عليه تقييم ذلك المفهوم البنائي بطريقة مماثلة أو متسـاوية في مناسبات أو حالات منفصلة. وبعبارة أخرى، فإن الدرجة الحقيقية للمتغير الكامن ينبغي أن تمارس تأثيراً مشابهاً أو مماثلاً على الدرجات الملاحظة في مناسـبتين (أو أكثر)، في حين ينبغي ألا يظل مكون الخطأ ثابتًا عبر تطبيقات المقياس. وبالتالي، ينبغي أن يمثل الارتباط بين الدرجات التي تم الحصول عليها عبر تطبيقـين للمقياس على نفس الأفراد المدى الذي يحدد به المتغير الكامن الدرجات الملاحظة. وهذا يعادل تعريف الثبات كنسبة التباين التي تعزى إلى الدرجة الحقيقية للمتغير الكامن.

إن مشكلة هذا المنطق تكمن في أن ما يحدث للدرجات بمرور الزمن قد يتعلق بتعرض إجـراءات القياس للخطأ أو لا يكون له صله بذلك. ويشـير «نونالي» (Nunnally, 1978)

إلى أن خصائص أو سـمات البنود قد تجعلها تنتج ردودًا مسـتقرة مؤقتًا حتى عندما يتغير المفهوم البنائي محل الاهتمام. على سـبيل المثال، إذا تأثر مقياس القلق المزعوم (المفترض) بالرغبـة الاجتماعيـة وكذلك القلق، فقـد تظل الدرجات ثابتة على الرغـم من الاختلافات أو التباينـات في القلق. لن يكون الاسـتقرار في الدرجات، الذي ينعكس في الارتباط المرتفع بين مناسـبات تطبيق المقياس، نتيجة الثبـات في الظاهرة موضع الاهتمام. وبدلاً من ذلك، قـد لا تتغير هـذه الظاهرة بينما تتغير الدرجات في المقيـاس، أي إن المقياس قد يكون غير ثابت. أو قد تُعزى التغييرات في الدرجات إلى عدم الثبات عندما تتغير الظاهرة نفسـها في الواقع، وقام المقياس بتعقب هذا التغيير بشـكل دقيق. والمشـكلة هي أن التغيير أو عدم التغيـير يمكن أن يُعزى إلى مجموعة متنوعة من الأشـياء إلى جانـب ثبات (أو عدم ثبات) إجـراءات القياس. حدد "كيلي وماكجراث" (Kelly and McGrath, 1988) أربعة عوامل مربكـة عندما يفحـص المرء مجموعتين من الدرجات لنفس المقيـاس، بينهما فاصل زمني. وهذه العوامل هي (١) التغيير الحقيقي في المفهوم البنائي محل الاهتمام (على سبيل المثال، صافي الزيادة في متوسـط مسـتوى القلق بين عينة من الأفراد)، (٢) التذبذبات أو التقلبات المنتظمة في الظاهرة (على سـبيل المثال، التباينات في القلق حول متوسـط ثابت كدالة في الزمـن)، (٣) التغييرات التي تعزى إلى الاختلافـات في الموضوعات أو طرق القياس بدلاً من الظاهـرة محل الاهتمام (مثل تأثيرات التعب والإجهاد الذي يتسـبب في حدوث أخطاء في قراءة البنود)، و(٤) عدم الاستقرار الزمني بسبب عدم الثبات المتأصل في إجراءات القياس. والعامل الرابع فقط هو عدم الثبات.

ومؤخـراً، ناقش "يـو" (Yu, 2005) كيف أن الأخطـاء في درجات ثبات إعـادة الاختبار لا تنشأ فقط من أوجه القصور في الأداة نفسها ولكن أيضاً تنتج من الأفراد الذين يتم تطبيق المقياس عليهم والإجراءات الخاصة بتطبيق الأداة. وتشـمل الأخطاء الناشـئة عن المشاركين الذيـن يجـري تقييمهم آثـار الانتقال التي يؤثر فيهـا الأداء في الاختبـار الأول على الأداء في الاختبار اللاحق. على سبيل المثال، قد يتذكر المجيب إجاباته السابقة وقد يرغب في أن يظهر بشكل متسق أو ثابت. وبالتالي، فإن الإجابات على التطبيق الثاني للاختبار لا تحددها بشكل مباشر حالة المتغير محل الاهتمام، بل يحددها الحافز على الظهور بشـكل متسـق مع مرور الزمن. يشـير «يو» إلى مصدر خطأ آخر يرجع إلى المجيب – وهو تأثير التشـاور أو المناقشة.

وكمثال، قد يأخذ طلاب المدارس الثانوية في البداية اختبار القدرة الدراسية أساساً كوسيلة لاكتســاب الخبرة في عملية الامتحان. ويمكنهم بعد ذلك التشــاور والمناقشــة بشــأن هذه التجربة واستخدامها كأساس للتحضير لاختبار لاحق. وكمثال أخر لتأثير التشاور أو التداول، يمكن للمشــاركين الذين يجدون أســئلة هجومية في التطبيق المبــدئي لأداة ما أن يتعمدوا حذف أو تشــويه إجاباتهم عندما يتم إعادة النظــر في البنود وقراءتها. إن أخطاء التطبيق التي ذكرها «يو» تتضمن اختلافات في الإجراءات بين الاختبار وإعادة الاختبار (على ســبيل المثال، الســماح للمســتجيبين بالعودة والتحقــق من الإجابات الســابقة)، والتعليمات غير الجيدة، والذاتية في تســجيل الإجابات (على سبيل المثال، المعايير الغامضة الخاصة بتسجيل الإجابات)، والفترة الزمنية غير المناســبة بين تطبيقات الاختبار. ويمكن لأي من هذه الأنواع من الأخطاء أن تقلل من درجات الثبات دون أن تكون الأداة نفسها غير ثابته بطبيعتها.

وهذا لا يعني أن إثبات الاستقرار الزمني أمر غير مهم. في أي عدد من سياقات البحث، قد يكون من الضروري افتراض (أو إثبات) أن القياســات المنفصلة أي التي توجد بينها فترة زمنيــة ترتبــط ارتباطًا وثيقًا (عالياً). ومع ذلك، فإن الاســتقرار الذي نســعى إليه في هذه الحالات هو استقرار كل من القياس والظاهرة. وتخبرنا ارتباطات إعادة الاختبار عن القياس فقط عندما نكون على ثقة كبيرة من أن الظاهرة ظلت مســتقرة ولم تتغير. وهذه الثقة لا يوجــد ما يبررها في كثير من الأحيان. وهكذا، فإن أفضل طريقة لثبات إعادة الاختبار، على الرغم من أهميتها، هي الكشف عن شيء ما حول طبيعة الظاهرة وقياسها، وليست الأخيرة (أي القيــاس) وحدها. وبالإشــارة إلى الثبات أو عدم التغــير في الدرجات بمرور الزمن فإن الاســتقرار الزمني هو الأفضل لأنه لا يشير، كما هو الحال بالنسبة لثبات إعادة الاختبار، إلى أن خطأ القياس هو مصدر أي عدم استقرار نلاحظه.

إن حساب ثبات إعادة الاختبار عادةً ما ينطوي على شكل ما من أشكال معامل الارتباط. والاختبــاران اللذان يفصل بينهما فترة زمنية يشــبهان ويماثــلان مجموعتين من البنود يتم تطبيقهــما في نفس الوقــت. وفي الحالة الأخيرة، وكــما ذكر آنفاً، فإن الارتبــاط بين البنود يعكس حاصل ضرب المســارين الممتدين من المتغير الكامــن إلى البنود. ووفقاً لافتراضات الاختبــارات المتكافئة التامة، تكون لهذه المســارات قيم متطابقة ولذلك فإن حاصل ضربها

تســاوي القيمة التربيعية لكل منها. كما يمثل كل مســار الارتباط بــين المتغير الكامن عند طرفــه الأصــلي والمؤشر عند طرفه النهــائي. ومن ثم، فإن مربع هــذه القيمة هو الارتباط التربيعي، أو نسبة التباين المشترك بين المتغير الكامن والمؤشر. لذلك، فإننا من خلال الربط بين البندين، نحصل على القيمة التي هي من الناحية المفاهيمية تعادل أو تساوي (في إطار افتراضــات الاختبار المتكافئ التام) نســبة التباين المشــترك بين البنــد والمتغير الكامن. وفي حالة ثبات إعادة الاختبار، يكون المنطق مطابقاً لذلك ولكن المؤشرين هما الآن الأداة التي تــم تطبيقها في الوقت الأول ثم تم تطبيقها مــره أخرى في الوقت الثاني. وينبغي أن يكون ارتباطهما هو تقديراً للارتباط التربيعي (وبالتالي، نسبة التباين المشترك) بين الدرجة التي تم الحصــول عليها (من أي من التطبيقين لــلأداة) والدرجة الحقيقية للمتغير الكامن. وعندما لا يتم الوفاء بافتراضات الاختبارات المتكافئة التامة، تصبح القيمة الناتجة حداً أدنى للثبات الحقيقي lower bound.

ويشــير هــذا التحليل إلى أنــه يمكن تقدير ثبــات مجموعه من البنود يتــم تطبيقها في مناســبتين أو فترتــين زمنيتــين من خلال الارتباط بــين الدرجات المتحصــل عليها من هذين التطبيقــين. وهذا، مــن الناحية العملية، هو طريقة حســاب ثبات إعــادة الاختبار عادة. في الظروف الشــائعة التي تكون فيها الدرجات الموجــودة في أداة الاختبار من نوع القياس المســتمر أو المتصل أساســاً، يتــم عادةً اســتخدام ارتباط بيرســون لحاصــل ضرب العزوم Pearson product-moment correlation. وينبغي أن يعمل ارتباط بيرسون بشكل جيد في معظــم الحــالات الخاصة بأدوات القياس التي تم تطويرها باســتخدام المنهج الموضح في هذا الكتاب. وإذا كان الباحث قلقاً من أن الدرجات المعنية لا تتطابق مع المقياس من نوع القياس الفتري أو المســافي (على سبيل المثال، القياس الترتيبي)، فإنه يمكن الاستعاضة بشكل بديل عن معامل الارتباط «بيرسون». ومع ذلك، لاحظ أنه حتى لو لم يكن كل بند من البنود على حدة من نوع القياس الفتري أو المسافي التام، فإن الدرجة القائمة على جمع هذه البنود معاً سوف تكون متقاربة إلى حد بعيد للبيانات المقاسة على مقياس فتري أو مسافي.

ثبات درجات التغيير:

هنــاك جانب آخر يســتحق منا الاهتمام عنــد النظر في الثبات عبر الزمــن. ونظراً لأن المقيــاس يظهر أنه يتمتع بالثبات، فإن هذا لا يتبعه بالضرورة أن درجات التغيير (المعروفة أيضًا باســم درجات الفرق أو الاختلاف)، والتي يتم حســابها على أنهــا الفرق بين درجات تطبيقــين لأداة القياس يفصل بينهما فترة زمنية، ســتتمتع بالثبات بنفس القدر. في الواقع، لاحظ العديد من المؤلفين (على سبيل المثال لين، سليندي (١٩٧٧)؛ أوفرال، وودوارد (١٩٧٥)، Linn & Slinde, 1977; Overall & Woodward, 1975) أن درجات التغيير تميل إلى أن تكون أقل ثبات من درجات مقياس تم تطبيقه مرة واحدة وتم الحصول على الدرجات من نفس أداة القياس.

وقبل النظر في هذا لموضوع ودراســته بشــكل أكبر، دعونا نفكر فيما تمثله ثبات درجة التغيــير. يفــترض مفهوم الثبات بالنســبة لدرجة التغيير أو الفــرق أن: (ا) هناك تباين في التغيير و(ب) أن جزءاً من هذا التباين يعزى إلى فروق حقيقية في التغيير بينما يُعزى الجزء الباقــي إلى عوامل أخرى (أي الخطــأ). ومن المفترض، أن الأفراد في موقف معين تكون لهم «درجــات حقيقيــة» مختلفة تمثل تغيراتهم الفردية بمرور الزمــن. أي إن القيمة الحقيقية المقابلة للتغيير في الزمن ســتختلف من فرد إلى فرد آخــر. إلى الحد الذي يتأثر فيه التباين بين الأفــراد في درجات التغيير بالاختلافات الفعلية في الدرجــات الحقيقية بدلاً من الخطأ، فــإن قيــاس التغيير هذا يكون ثابتاً. لنأخذ مثالاً مبالغاً فيه (متطرف)،إذا تغير كل شــخص بنفس المقدار تمامًا، فلن توجد اختلافات فردية في الميل إلى التغيير. أي إنه لا يوجد اختلاف بــين الأفراد في التغيير الحقيقي، وبالتالي، لا يوجد أي ارتبــاط بين التغيير الحقيقي والتغيير الملحــوظ (كما هو الحال دائمًا عندما يكون أحدهم ثابتًــا). وبالتالي، يجب أن تكون قيمة ثبــات درجة التغيير (أي الارتبــاط التربيعي بين الدرجات الحقيقيــة والدرجات الملاحظة) تســاوي الصفر أيضاً. وبديهياً، إذا تغير الجميع في انسجام تام، فمن المحتمل أن نفترض أن التغيير يمثل نوعًا من التأثير الظرفي بدلاً من السمات الشخصية البحتة. إذا كان الشيء الذي يهمنا قياسه هو سمة شخصية، فإن درجة التغيير التي تعكس في المقام الأول ظاهرة موقفية وليست الظاهرة الشخصية هدفنا.

كيف تبدو معادلة ثبات درجات التغيير؟ وكما هو الحال مع أي معامل ثبات، فإن نسبة التباين الكلي يمكن أن تعزى إلى التباين الحقيقي. وهكذا:

$$r_{\Delta} = \frac{\sigma^2_{\Delta T}}{\sigma^2_{\Delta O}} = \frac{\text{تباين الدرجة الحقيقية}}{\text{تباين الدرجة الملاحظة}}$$

حيث (r_{Δ}) هي ثبات درجة التغير، و($\sigma^2{}_{\Delta T}$) هي تباين درجة التغيير الحقيقية، و($\sigma^2{}_{\Delta 0}$) هـــي تباين درجة التغيير الملاحظ (أي مجموع تباينات الدرجــة الحقيقية والخطأ). ويمكننا التعامــل مع تباينات درجــات التغير كأن لدينــا تباينات أخرى من التوليفات، أي تســاوي مجموع كل تباينات المكونات، بالإضافة إلى ضعف تغايراتها. وهذا مماثل لتوسيع النطاق أو المقدار الجبري $(X + Y)^2$ إلى $X^2 + Y^2 + 2XY$.

ومع ذلك، نظراً لأننا نتعامل مع درجات الفروق على الشــكل $D = X2 - Y1$ فإننا نقوم بعملية طرح وليســت عملية إضافة أو جمع، لذلك فإن المقدار الجبري هو في الواقع مماثل للتوسيع $(X + Y)^2$ into $X^2 + Y^2 - 2XY$. ولأن $(X2 - X1)^2 = X1^2 + X2^2 - 2\ (X1 * X2)$، عندما نوسع التباين لدرجة الفرق أو التغيير، فإننا نصل إلى:

$$r_{\Delta} = \frac{\sigma^2_{T_{X1}} + \sigma^2_{T_{X2}} - 2\sigma_{X1,X2}}{\sigma^2_{X1} + \sigma^2_{X2} - 2\sigma_{X1,X2}}$$

حيث يكون المقدران الأوليان في البســط هما تباينات الدرجات الحقيقية لأداة القياس الأوليــة وأداة القياس اللاحقة، عــلى التوالي، للمقياس موضع الاهتــمام؛ والمقدار الأخير في البســط هو تغايراتهم. تتوافــق مقادير المقام مع نظيراتها في البســط، لكن التباين يكون للدرجات الملاحظة وليس للدرجات الحقيقية.

يمكننــا تقدير تباينات الدرجــات الحقيقية من خلال حاصــل ضرب ثبات المقاييس في تباينات درجاتها الملاحظة. وهذا، في جوهره، يعطينا نســبة تباين الدرجات الملاحظة التي نعتبرها ناتجة أو ناشــئة مــن الدرجات الحقيقية بدلاً من الخطأ. باســتخدام هذا الإجراء لتقدير التباين الدرجات الحقيقة، يكون المقدار السابق مساوياً لـ:

$$r_{\Delta} = \frac{\left[\left(\sigma^2_{X1}\right)\times\left(r_{X1,X1}\right)\right]+\left[\left(\sigma^2_{X2}\right)\times\left(r_{X2,X2}\right)\right]-\left[2\sigma_{X1,X2}\right]}{\left(\sigma^2_{X1}\right)+\left(\sigma^2_{X2}\right)-\left(2\sigma_{X1,X2}\right)},$$

حيــث ($r_{x1,x1}$) و($r_{x2,x2}$) هي ثبات القياســات الأولى والثانية على التوالي. لاحظ أنه في كلتا المعادلتين السابقتين، يكون التغاير (التباين المشترك) متطابقًا في كل من البسط والمقام. ويرجع ذلك إلى أنه إذا افترضنا أن الأخطاء غير مرتبطة عبر الزمن، فإن كل التغاير عبر الزمن هو تباين يتميز بالثبات. وهذا يعني أن التغاير في الدرجات الملاحظة لا يمكن أن ينشــأ إلا عن التغاير للدرجات الحقيقية إذا اســتمر افتراض عدم ارتباط الأخطاء قائماً. وبالتالي، فإن التغاير الملاحظ والتغاير الحقيقي يكونان نفس الشيء (متساويان).

باســتخدام المعادلة السابقة، يمكننا عمل جدول لماهية ثبات درجة التغيير التي ستكون لقيم مختلفة لتباينات الدرجات والتغايرات. لنفترض أن لكلا الدرجات الناتجة عن تطبيق الأداة في الوقــت الأول والوقت الثــاني تباين الوحدة وأن للمقيــاس موضع الاهتمام نفس الثبات في كل من الوقت الأول والوقت الثاني. يمكننا الآن أن نرى كيف يؤثر التغاير الخاص بهما على ثبات درجة التغيير.

	ثبات درجة التغيير ($r\Delta$) عندما تكون درجة الثبات		
$\sigma_{X1,X2}$	$r_{X1,X1} = r_{X2,X2} = .50$	$r_{X1,X1} = r_{X2,X2} = .70$	$r_{X1,X1} = r_{X2,X2} = .90$
.00	.50	.70	.90
.10	.44	.67	.89
.20	.38	.63	.88
.30	.29	.57	.86
.40	.17	.50	.83
.50	.00	.40	.80
.60	—	.25	.75
.70	—	.00	.67
.80	—	—	.50
.90	—	—	.00

مـا نراه هـو أن الثبات الأعلى في وقت واحد يرتبط بزيـادة ثبات درجة التغيير (أي إن الزيـادة في قيم ثبات درجة التغيير من اليسـار إلى اليمين عبر أي صف من الجدول)، وهو أمر غير مفاجئ أو مثير للدهشة. ما قد يكون أكثر إثارة للدهشة في البداية هو أن التغاير الأعـلى للدرجات عبر الزمن يـؤدي إلى انخفاض ثبات درجة التغيير (أي انخفاض قيم ثبات درجة التغيير من أعلى إلى أسفل على طول أي عمود من الجدول).

يشـير روجوسـا (Rogosa (1988، الذي كان من بين أقوى المؤيديـن لدرجات التغيير، إلى أن الجـداول التي تعطي قيم الثبات فقط للدرجات التي ترتبط ارتباطًا كبيرًا عبر الزمن (مثل تلك الموجودة في الصفوف السفلية من الجدول السابق) قد ترسم صورة قاتمة بشكل لا مـبرر لها لثبات درجـات التغيير. ويجادل «روجوسـا» بأن التغاير العـالي في الدرجات عبر الزمن يقيد أو يحد من نسبة التباين الملاحظ في الوقت الثاني والذي يمكن أن يعزى إلى الاختلافات الفردية في التغيير. والسبب في التغاير المرتفع مع مرور الزمن هو أن كل شخص يتغير بشـكل أسـاسي في انسجام تام بدلاً من أن يكون هناك تباين كبير بين شخص وشخص آخـر في تغيـير الدرجات بمرور الزمـن. إذا كان هناك تباين أو اختلاف بسـيط في المعدل الحقيقـي للتغيـير، فلا يمكن أن يكون هناك الكثير من التغايـر بين درجة التغيير الحقيقية ودرجـة التغيير الملاحظة (تذكر أنه لكي يكون هناك تغاير كبير يجب أن يكون هناك تباين كبـير لكلا المتغيريـن). ويذهب «روجوسـا» (١٩٨٨) إلى القول بأنـه، عندما يكون المدى الحقيقي للتباين في معدل التغير كبيرًا، فإن ثبات درجات التغيير بالنسـبة لثبات الدرجات في نقطـة زمنية واحدة تكون مواتية بدرجة كبيرة. توضح الصفوف العليا من جدولنا هذه النقطة. ومع ذلك، فإن الجداول المنشورة في كثير من الأحيان لم تقدم إلا الظروف القصوى المبينة في الصفوف السفلية من الجدول السابق.

ماذا يعني كل هذا على المستوى العملي أو التطبيقي؟ من المتوقع أن تؤدي الظروف التي تحدث فيها الأحداث على مر الزمن إلى تأثير مماثل من حيث الحجم على جميع المشاركين، مما سيؤدي إلى ضعف ثبات درجات التغيير لأنه لن يكون هناك تباين يذكر بين الناس. وهذا يعني أنه إذا كانت حصة الأسـد (الحصة الأكبر) من التباين ناتجة عن عوامل تُعزى إلى مرور الزمن (مثل تلقي تدخل يؤثر على جميع المشاركين بالتساوي)، فإن الحصة التي تُعزى إلى الاختلافات الفردية يجب أن تكون صغيرة نسـبياً في المقابل. إذا زادت درجات كل الأشـخاص من الوقت

الأول إلى الوقت الثاني بمقدار خمس نقاط، على سبيل المثال، فيمكنك أن تتنبأ أو تتوقع بدقة درجة الوقت الثاني من خلال الدرجة السابقة. وسيكون ارتباط الدرجات في الوقتين هو واحد صحيح، ودرجة التغيير الخام ستكون ثابتة (أي خمس نقاط لكل الأشخاص)، وبالتالي فإن ثبات درجة التغيير ستكون صفراً. وعلى العكس من ذلك، فإن الحالات التي يرجح أن تختلف فيها التغــيرات بمرور الزمن اختلافاً كبيراً من فرد إلى آخر تخلق على الأقل فرصة لكي تكون درجات التغيير أكثر ثباتاً. وإذا نظرنا إليها من منظور مختلف قليلاً، إذا كان مقدار التباين قد ساهم في درجــات الوقت الثاني من خلال درجات الوقت الأول يبلغ ١٠٠٪ تقريباً، فإن مقدار التباين في درجة الوقت الثاني لا يمكن أن تنسب إلى متغيرات أخرى غير درجة الوقت الأول. يتم تحديد درجة الوقت الثاني أساساً من خلال الدرجات التي تم تسجيلها، بغض النظر عما قد حدث بين الوقــت الأول والوقت الثاني. ومن ثم، فإن مؤشر التغيير المفترض في متغير معين غير حســاس للتغيرات التي تحدث مع مرور الزمن لا يؤدي وظيفته المقصودة.

كما تناول «تراوب» Traub (1994) أيضاً مسألة ثبات درجات التغيير أو الفرق. ويقترح أن الافتراضــات حــول المقاييس المعنية قد تكون إشــكالية (معضلة) بشــكل خاص عندما يخطــط المرء للنظر في درجات التغيير الخام (على ســبيل المثال، درجــات الاختبار البعدي مطروحــاً منها درجة الاختبار القبلي أو التمهيــدي). وعلى وجه التحديد، يلاحظ "تراوب" أن اتســاق ما يقيســه المقياس داخل الفترات الزمنية وعبرها يكتســب أهمية خاصة عند إدراج وتضمــين درجــات التغيير. على ســبيل المثال، قد يمثل الخروج عن القياس المســافي أو الفــتري الحقيقــي، حتى خلال فترة زمنية واحدة، مشــكلة لأن التغير في بعض أجزاء من نطاق الدرجات قد لا تكون قابلة للمقارنة من حيث الحجم بالنسبة إلى التغيرات المتطابقة عدديــاً في نقاط مختلفة على طــول المقياس. وبالتالي، فإن التغير من درجة (٢) إلى (٣) قد لا يعــادل التغــير من درجة (٥) إلى (٦) على مقياس الإعاقــة البدنية أو العجز البدني الذي يــتراوح بين درجة واحدة إلى عشر درجات. وثمة مثال آخر يتعلق بالتغيرات من فترة زمنية إلى أخرى. قد تتغير الاختلافات (الفروق) فيما يقيســه المقياس نفســه مرتين، إذا استخدم الأشخاص، على ســبيل المثال، ذاكرة إجابتهم السابقة كدليل لتحديد إجاباتهم على الاختبار البعدي بدلاً من أو بالإضافة إلى السمة التي يراد أن يقيسها المقياس. كما يلاحظ «تراوب» أيضاً أن المحاولات الرامية إلى تقليل هذه المشــكلة إلى أدنى حد باستخدام مقاييس مختلفة

(مثــل النماذج المتكافئة أو المتوازية) في الوقــت الأول والوقت الثاني قد تؤدي إلى حدوث مشــكلات إضافية. واحدة من هذه المشــكلات هي المعايرة المختلفة أو التدريج المختلف. ما لم تكن وحدات التغيير على المقياس الأول تعني بالضبط نفس الشيء الذي تعنيه وحدات التغيير في المقياس اللاحق، يمكن أن تكون درجات التغيير مضللة. حتى في حالة اســتخدام الاختبارات المتكافئة المزعومة، يحذر «تراوب» من وجود اختلافات طفيفة في المعايرة أو في المتغيرات الكامنة التي تحدد درجاتها. وبالتالي، فإنه ينصح بعدم استخدام درجات التغيير.

ويصف «هيويت، وأندرســن، ومينــور» (1992) Hewett, Anderson, and Minor خمس إســتراتيجيات تحليلية بديله لتصاميم البحوث التــي تنطوي على التغيير. بالإضافة إلى (أ) استخدام درجات التغيير كمتغيرات تابعة، تتضمن هذه الطرق (ب) اختبار التفاعل بين عضوية المجموعة والوقت، (ج) اختبار الاختلافات أو الفروق في المتوســطات في الوقت الثاني، (د) اختبار الاختلافات أو الفروق في متوسطات الوقت الثاني بعد التحكم في الدرجات في الوقــت الأول عــن طريق تحليل التبايــن المصاحب أو المشــترك (ANCOVA)، و(هـ) باستخدام تحليل التباين المصاحب أو المشــترك ANCOVA اللامعلمي nonparametric للضبط والتحكم في متوسطات الوقت الأول. ويخلص المؤلفون إلى أنه من الضروري مراعاة الدرجــات الأولية إذا كانت هذه الدرجات مرتبطــة بدرجات لاحقة (كما هو الحال غالباً). وبالتالي، فإنهم يوصون باســتخدام حلول تحليل التباين المصاحب أو المشترك ANCOVA. والمنطق هو أن تأخذ في الاعتبار الحالة الأولية للمتغير التابع قبل فحص تأثير المتغير المستقل. يمكــن تنظيم تحليل لهذا النوع بطريقتين متماثلتين عــلى الأقل. أحدهما هي التعامل مع المســتوى الأوَّلي للمتغــير التابع على أنه متغير مصاحب covariate والتعامل مع مســتواه في وقــت لاحق كنتيجة. وعلى نحوٍ متســاوٍ، يمكن للمرء اســتبدال درجــة التغيير كنتيجة. هاتان الطريقتان متكافئتان من الناحية الرياضية لأنه من بين الدرجات الأولية، والدرجات اللاحقة، ودرجات التغيير، لا يوجد في الواقع سوى جزأين مستقلين من المعلومات. ومعرفة أي اثنــين منهما يحدد الثالث (أي، هناك درجتان فقط مــن الحرية بين الدرجات الثلاث). التحكم في الحالة الأولية للمتغير هو جزء واحد من المعلومات. ويمكن تمثيل الجزء المتبقي باســتخدام أي من الجزأيــن المتبقيين كنتيجة. ومع ذلك، من المهم أن ندرك أن اســتخدام درجــة التغيير كنتيجة عندما يتم التحكم في الحالة الأولية لا يختلف عن اســتخدام درجة

التغيــير الخام كنتيجة دون تعديل للحالــة الأولية للمتغير. إن إدراج الحالة الأولية كمتغير ضابط أو متغير تحكم يتجنب العديد من المشاكل المرتبطة بعدم ثبات درجة التغيير بسبب أنها درجة التغيير الخام (أي غير المعدلة) التي تخضع لعدم الثبات.

وفي الآونــة الأخــيرة، لاحــظ «كرويــن وإيمونــز وسيجتســما» Kruyen, Emons, and Sijtsma (2014) مشــكلات أو قضايا إضافية يمكن أن تصيب أو تكون مصدر إزعاج لدرجات التغيير. وبشــكل أكثر تحديدًا، تتناول ورقتهم الاستخدام الشــائع لدرجات التغيير، أي تقييم التغيير الفردي في الســياق السريري. على سبيل المثال، قد يرغب أحد ممارسي الصحة العقلية في تقييم ما إذا كان المريض قد عانى من أعراض منخفضة أو أقل بعد تلقي العلاج أو ما إذا كان هذا المريض قد تحول وانتقل عبر حد التشخيص وأنه لم يعد من الممكن اعتباره مبررًا لتشخيص ســابق. قد لا يتم استخدام ثبات المقياس على نحو مناسب لتحديد ما إذا كانت الاستنتاجات على المستوى الفردي مبررة أم لا. وكما يلاحظ «كروين وآخرون» (Kruyen et al, 2014) أن الثبات هو في الأســاس خاصية لأداء المقياس عبر مجموعة من المجيبين. وبالتالي، فإن تطبيقه لإصدار أحكام بشــأن التغييرات في الأفراد يثير مشــاكل. ينشــأ أحد الســيناريوهات الشائعة ولكن إشــكالي بشكل خاص عندما يحاول الطبيب تحديد مقدار تغير الفرد بناءً على اختلافات الدرجات بمرور الزمن باســتخدام مقياس قصير نســبيًا. ويحــذر كروين وآخرون (٢٠١٤) من أنــه في الحالات التي يكون فيها المقياس مكوناً من أقل من عشرة بنود، فإن اســتخدامه بهذه الطريقة يمكن أن يؤدي إلى معدل مرتفع بشكل غير مقبول من الأحكام الخاطئة.

وكما هو الحال بالنســبة للعديد مــن الموضوعات الأخرى في هــذا الكتاب، فإن أفضل الممارســات التي ينبغي اتباعها هي التفكير في الظروف المتاحة وكيفية تعاملها مع البيانات المتاحة والافتراضات الخاصة بمختلف النهج أو الطرق. ستكون هناك ظروف تجعل درجات التغيير ذات مغزى كبير، ولا ســيما عندما تؤخذ الدرجات الأولية في الاعتبار، تعمل بشــكل جيد. وفي حالات أخرى، ولا سيما عند التفكير في استخدام درجات التغيير الخام، قد يكون هنــاك ما يبرر اتباع نهج آخر. إن المجموعة الكاملة مــن الظروف الممكنة والنهج الممكنة كثيرة للغاية بحيث لا يمكن مناقشــتها هنا. ومع ذلك، لا ينبغي تجاهل الرســالة الأساسية لهذا القسم - وهي أن ثبات درجة التغيير لا يمكن اعتبارها مساوية لثبات الدرجات لمقياس يتم تطبيقه في وقت واحد.

الثبات والقوة الإحصائية:

مــن الفوائد التي كثيراً ما يتــم تجاهلها أو التغاضي عنها في كثير من الأحيان في المقاييس الأكــثر ثباتاً أنها تزيد من القوة الإحصائية لحجم عينة معين (أو تســمح لحجم عينة أصغر بتوليد قوة مكافئة أو مســاوية لحجم عينة أكبر)، مقارنــةً بالمقاييس الأقل ثباتاً. وللحصول عــلى درجــة محددة من الثقــة في القدرة على اكتشــاف فرق بمقدار معــين أو معنوي بين مجموعتين تجريبيتين، على سبيل المثال، يحتاج المرء إلى عينة ذات حجم معين. ويمكن زيادة احتمال اكتشــاف مثل هذا الاختلاف أو الفرق (أي قوة الاختبار الإحصائي) عن طريق زيادة حجم العينة. وفي العديد من التطبيقات، يمكن الحصول على نفس التأثير عن طريق تحسين ثبــات المقيــاس. والمقياس الذي يتمتع بالثبات، مثل العينة الكبيرة، يســهم في تقليل الخطأ نســبياً في التحليل الإحصائي. قد يكون من الأفضل للباحثين أن يقوموا بتقييم المزايا النسبية لزيادة ثبات المقياس مقابل حجم العينة في حالات البحث التي يتوفر فيها كلا الخيارين.

تعتمد القوة المكتســبة من تحســين الثبــات على عدد من العوامــل، بما في ذلك حجم العينة الأولية، ومســتوى الاحتمال المحدد لاكتشــاف خطأ من النوع الأول Type I error، وحجم التأثير (مثل، متوســط الفرق) الذي يعتبر معنوياً أو ذا دلالة إحصائية، ونسبة تباين الخطــأ الذي يُعزى إلى قياس عــدم الثبات بدلاً من عدم تجانــس العينة أو مصادر أخرى. وتتطلــب المقارنة الدقيقة بين تحســين الثبات وزيادة حجم العينــة تحديد هذه العوامل؛ ومــع ذلك، توضح الأمثلة التالية هذه النقطــة. في حالة بحث افتراضية يكون فيها احتمال وجــود خطأ من النــوع الأول (٠,٠١)، والفرق يبلغ عشر نقاط بين المتوســطين يعتبر مهماً، وتباين الخطأ يســاوي ١٠٠، فإنه يجب زيادة حجم العينة من ١٢٨ إلى ١٧٢ (زيادة العينة بنســبة ٣٤ ٪) لرفع قــوة اختبار (ف) F من ٠,٨٠إلى ٠,٩٠وســيكون لتقليــل تباين الخطأ الإجمالي من ١٠٠ إلى ٧٥ (انخفاض بنســبة ٢٥ ٪) نفس النتيجة بشــكل أساسي دون زيادة حجم العينة. وقد يحقق اســتبدال مقياس أقل ثباتاً بمقياس آخر يتمتع بمستوى ثبات عالٍ نفس النتيجة. وكمثال آخر، بالنسبة لـعينة حجمها يساوي ٥٠ ($N = 50$)، هناك مقياسان بمســتوى ثبات قدره ٠,٣٨ ويرتبطان معاً بمعامل ارتباط قــدره ($r = 0.24$) بالكاد يحققان معنوية أو مستوى دلالة إحصائية عند ($p <.10$). إذا تمت زيادة مستوى ثباتهما إلى ٠,٩٠.

إذا بقي الثبات عند ٠,٣٨، فستكون هناك حاجة إلى مضاعفة حجم العينة من أجل الوصول إلى معامل ارتباط ذي دلالة إحصائية عند (p <.01). ويوفر لبيسي Lipsey (١٩٩٠) مناقشة أكثر شمولاً حول القوة الإحصائية، بما في ذلك تأثيرات ثبات القياس.

يتم تعزيز وتحسين الثبات إما من خلال عدد البنود أو الفقرات ومتوسط الارتباط بين البنود (والذي ينشأ من ارتباط أقوى بين كل بند والدرجة الحقيقية)؛ إما إضافة المزيد من البنود أو الحصول على بنود أفضل يزيد من الثبات للمقياس. ويمكن لأي من طريقتي تحسين الثبات أن تزيد وتعزز القوة الإحصائية. ومرة أخرى، هذا يشبه ويماثل الوسيلة التي يمكن بها للمعاينة (طرق أخذ العينات) أن تزيد من القوة الإحصائية. ويمكن أن يؤدي تقليل خطأ المعاينة، إما عن طريق اختيار عينات «أكبر من حيث الحجم» (أي، جمع البيانات من عينة أكبر) أو أخذ عينات «أكثر ذكاءً» (أي، جمع البيانات من الأفراد الذين يمثلون المجتمع الإحصائي موضع الاهتمام أفضل تمثيل)، إلى زيادة القوة الإحصائية. وبالمثل، يمكن أن تقلل المقاييس الأكبر (التي تتضمن بنوداً أكثر) والأكثر ذكاءً (وجود ارتباطات أقوى بين البنود وبعضها البعض) من التباين الناتج عن خطأ القياس، وبالتالي، زيادة القوة الإحصائية.

نظرية التعميم:

حتى الآن، لقد ركزت مناقشتنا حول الثبات على تقسيم التباين الملاحظ إلى الجزء الذي يُعزى إلى الدرجة الحقيقية للمتغير الكامن والجزء المتبقي، وهو الخطأ. يقدم هذا القسم بإيجاز إطار عمل أكثر عمومية لتقسيم التباين بين مصادر الخطأ والمصادر الأخرى غير الخطأ. ويعود هذا الإطار إلى مناقشتنا السابقة للنُهج أو المداخل القائمة على تحليل التباين ANOVA للتمييز بين مصادر التباين.

قبل أن نطبق فكرة التقسيم الأكثر دقة لتباين الخطأ على القياس، دعونا ننظر في مثال بحثي أكثر عمومية يتم فيه فحص مصادر التباين المتعددة. لنفترض أن الباحث أراد تحديد فعالية برنامج تدريبي يهدف إلى زيادة الإنتاجية المهنية. علاوة على ذلك، نفترض أن الباحث قام بتطبيق وإدارة البرنامج التدريبي لعينة كبيرة من أساتذة الجامعات وعلى عينة مماثلة من الفنانين. وحدد الباحث أيضًا مجموعات مماثلة من الأساتذة والفنانين الذين لن

يشاركوا في البرنامج التدريبي ولكنهم سيشاركون في نفس تقييم الإنتاجية الذي يشارك فيه المشاركون في البرنامج التدريبي. عند التفكير في الدراسة بعض الشيء، قد يكون هذا الباحث خلُص إلى أن ملاحظات الإنتاجية ستعكس تشغيل ثلاثة مصادر محددة للتباين المنهجي أو المنتظم: (١) مشــارك مقابل غير مشارك، (٢) أستاذ جامعي مقابل فنان، و(٣) تفاعل هذه الآثــار أو التأثيرات. وتتمثل إحدى الإســتراتيجيات التحليليــة المعقولة أو المنطقية في هذا الموقــف في إجراء تحليل التباين ANOVA على درجات الإنتاجية، ومعالجة كل مصدر من مصادر التباين هذه كبعد في التحليل. ويمكن للباحث عندئذ أن يحدد إلى أي مدى يساهم كل مصــدر مــن مصادر التباين في التباين الكلي في الإنتاجية المهنية. في جوهرها، فإن هذه الإستراتيجية التحليلية ستقسم التباين الكلي بين درجات الإنتاجية الملاحظة إلى عدة مصادر هي: المشــاركة في التدريب، والمهنة، وتفاعل هــذه العوامل، والخطأ. ويمثل الخطأ جميع مصادر التباين بخلاف تلك المحددة بواسطة العوامل السابقة.

الآن، لنأخــذ في الاعتبــار حالة افتراضية يقــوم فيها الباحث بتطويــر مقياس الرغبة في الاســتقلال الذاتي. وســوف يســتخدم هذا المقياس في دراسة للمســنين، والذين قد يواجه بعضهــم مشــاكل بصرية. وبالتــالي، يخطط الباحــث لإدارة أو تطبيق مقيــاس الرغبة في الاســتقلال الذاتي شفوياً لهؤلاء لأشخاص الذين ســيصعب عليهم القراءة وفي شكل مكتوب للمشاركين المتبقين في الدراسة.

إذا تجاهل الباحث طريقة إدارة أو تطبيق المقياس (المكتوبة مقابل الشــفهية) كمصدر للتباين في درجات الاختبار، فإنه ســيكون فيما يتعلق بكل درجة تم الحصول عليها بسبب المســتوى الحقيقــي لرغبة المجيب في الاســتقلال الذاتي بالإضافة إلى درجــة ما من الخطأ. ويمكن للباحث أن يواصل حســاب الثبات كما ســبق مناقشته. ومع ذلك، لاحظ أن مجرد حســاب معامل "ألفا" على درجات المقيــاس دون الأخذ في الاعتبار طريقة إدارة أو تطبيق المقيــاس لن يميز الخطأ المنهجي المحتمــل الناتج عن طريقة تطبيق المقياس عن أي مصدر خطأ آخر.

وكبديــل لذلك، يمكن للباحث الاعتراف بوضــع طريقة إدارة أو تطبيق المقياس كمصدر للتباين بين الدرجات، باستخدام أسلوب تحليل التباين. إذا أظهرت نتائج التحليل أن الفرق

بين أساليب إدارة وتطبيق المقياس يمثل نسبة غير مهمة من التباين الكلي في الدرجات، حينئذ يمكن للباحث أن يكون لديه ثقة أكبر في إمكانية مقارنة الدرجات بالنسبة للأفراد الذين أكملوا الإجابة على المقياس إما بالصيغة الشفوية أو المكتوبة. أما إذا كان هناك مقدار كبير من التباين الكلي الملاحظ في الدرجات يُعزى إلى طريقة إدارة وتطبيق المقياس، فإن الباحث سيعلم أن أي تفسير للدرجات ينبغي أن يأخذ هذا الفرق بين الأنماط في الاعتبار.

توفر نظرية التعميم (على سبيل المثال، Cronbach, Gleser, Nanda, & Rajaratnam, 1972) إطارًا لدراسة مدى إمكانية افتراض تكافؤ عملية القياس عبر بُعد واحد أو أكثر. وفي المثال السابق، كان البعد موضع الاهتمام هو طريقة إدارة وتطبيق المقياس. وكل بُعد من الأبعاد موضع الاهتمام هو مصدر محتمل للتباين ويشار إليه على أنه وجه واحد أو جانب واحد facet. ويركز المثال على أسلوب إدارة وتطبيق المقياس باعتباره المصدر الوحيد المحتمل للتباين (بخلاف الأفراد) الذي يرغب الباحث في تعميمه. لذلك، هذا المثال ينطوي على وجه أو جانب واحد.

وفي لغة نظرية التعميم، تشكل الملاحظات التي يمكن الحصول عليها عبر جميع مستويات أحد الأوجه (على سبيل المثال، مع كل من أسلوب الإدارة أو التطبيق الشفوي والمكتوب للمقياس) عالمًا من الملاحظات المقبولة. ويشار إلى المتوسط الحسابي لهذه الملاحظات باسم الدرجة الكلية universe score وهي تشبه الدرجة الحقيقية في نظرية الاختبار الكلاسيكية classical test theory (Allen & Yen, 1979). وتسمى الدراسة التي تهدف إلى تحديد مدى قابلية مقارنة الدرجات عبر المستويات المختلفة من الوجه باسم "دراسة إمكانية التعميم generalizability study" أو «G-study». وتعد الدراسة الافتراضية للرغبة في الاستقلال الذاتي مثالًا على دراسة "إمكانية التعميم" بحكم معالجتها لتأثيرات المستويات المختلفة لوجه أو جانب طريقة إدارة أو تطبيق المقياس.

والغرض من دراسة إمكانية التعميم «G-study» هو مساعدة الباحث في تحديد المدى الذي يقوم به الوجه facet في الحد من التعميم أو عدمه. إذا كان هناك وجه (على سبيل المثال، أسلوب الإدارة أو التطبيق) يفسر قدرًا كبيرًا من التباين في الدرجات الملاحظة، فإن النتائج لا تعمم عبر المستويات (مثل الإدارة الشفوية مقابل الإدارة المكتوبة) لهذا الوجه أو الجانب.

ويتم التعبير عن مدى إمكانية التعميم عبر مســتويات الوجه facet دون تحريف البيانات بمعامل إمكانية التعميم generalizability coefficient. ويتم حســاب هذا المعامل عادةً عن طريق تكوين نسبة من متوسط المربعات المناسبة الناتجة عن تحليل التباين التي يتم إجراؤها كجزء من دراســة إمكانية التعميــم G-study. ومن الناحية النظرية، فإن معامل إمكانية التعميم هو نســبة التباين في الدرجات الكلية إلى تباين الدرجات الملاحظة وبالتالي فهو مماثل ومشــابه لمعامــل الثبــات (Allen & Yen, 1979). ومع ذلــك، لاحظ أنه إذا أسفرت دراســة إمكانية التعميم عن معامل تعميم ضعيف، فإن تصميم الدراسة يشير إلى مصدر المشــكلة (أي، الجانب أو الوجه الذي تم فحصة). فمعامل الثبات لا يحدد ســوى مقدار الخطأ دون أن يعزى ذلك إلى مصدر محدد.

وفي بعض الحالات، قد يكون اختيار تصميم تحليل التباين المناســب، وتحديد التأثيرات التــي تتوافــق أو تتطابق مــع جوانب الاهتمام، وبنــاء معامل إمكانيــة التعميم الصحيح أمــراً صعباً. كما هو الحال مع تحليل التباين بشــكل عام، يمكــن لأبعاد متعددة، وتأثيرات متداخلــة، ومتقاطعة، ومختلطة أن تجعل دراســة إمكانية التعميــم معقدة. (انظر مايرز ١٩٧٩؛ أو كــيرك ١٩٩٥، Myers 1979, or Kirk 1995، للاطــلاع عــلى المناقشــات العامة حــول تصميمات تحليــل التباين) وينصــح بالحفاظ على تصميم دراســة إمكانية التعميم بســيطة. ومن الحكمة أيضاً الرجوع إلى مصدر يــشرح بالتفصيل كيفية بناء نموذج تحليل التبايــن المناســب لنوع معين من دراســات إمكانيــة التعميم. ويصف «كروكــر والجينا» (Crocker and Algina,1986) التصميــمات المناســبة للعديد من دراســات مختلفة لها جانب أو وجه واحد واثنين من دراســات إمكانيــة التعميم. كما يوفر هذا المصدر مقدمة عامة جيدة لنظرية التعميم.

ملخص:

تتمتع المقاييس بالثبات بقدر ما تتكون من بنود أو فقرات ثابته تشــترك في متغير كامن عام أو مشــترك. ويتوافق معامل «ألفا» بشــكل وثيق مع التعريف التقليدي للثبات كنسبة التبايــن في المقيــاس الذي يعــزى إلى الدرجة الحقيقيــة للمتغير الكامــن. الطرق المختلفة لحســاب الثبات لهــا فائدة مختلفة في حالات معينة. على ســبيل المثــال، إذا لم يكن لدى

الشخص إمكانية الوصول إلى إصدارات متكافئة من مقياس معين، فإنه من المستحيل حساب ثبات النماذج البديلة. الباحث الذي يفهم مزايا وعيوب الأســاليب البديلة لحســاب الثبات يكــون في وضع أفضل لاتخاذ أحكام مســتنيرة عند تصميم دراســة القيــاس أو تقييم تقرير (دراســة) منشــور. إن الثبات هو قضية تتجاوز الاهتمامات والســياقات البحثية المختلفة. وتستخدم كلا من نظرية إمكانية التعميم ونهج معامل الارتباط داخل الطبقة لتقييم الثبات وكلتا الطريقتين تســتغل قدرة تحليل التباين لعزل مصادر التباين المنفصلة وقياســها كمياً. وهكذا، فإن الأنشطة البحثية التي تبدو على أنها متباينة كدراسة لإمكانية التعميم، والتقييم النفسي(السيكومتري) لبنود الاستبانة، وتقييم ثبات المقيّمين. جميعها تقوم على تعريف عام ومشترك للثبات كنسبة التباين التي تعزى إلى الدرجة الحقيقية للظاهرة موضع الاهتمام.

تمارين:

١- إذا كانــت هناك مجموعة من البنود أو الفقرات لديها اتســاق داخلي جيد، ماذا يعني ذلك بشأن علاقة هذه البنود بالمتغير الكامن الخاص بها؟

٢- في هذا التمرين[٣]، افترض أن ما يلي هو مصفوفة تغاير لمقياس (Y) يتكون من ثلاثة بنود هي (X1 وX2 وX3):

٠,٤٠	٠,٥٠	١,٢
٠,٦٠	١,٠	٠,٥٠
١,٨	٠,٦٠	٠,٤٠

أ - ما هو التباين لكل من البنود الثلاثة (X1 وX2 وX3)؟

ب- ما هو تباين المقياس (Y)؟

جـ- ما قيمة معامل «ألفا» للمقياس (Y)؟

٣- ناقش الطرق التي يربك بها أو يخلط أسلوب ثبات إعادة الاختبار test-retest reliability عوامل أخرى بخصائص المقياس الفعلي.

٤- كيف ينبثق منطق ثبات النماذج البديلة من افتراضات الاختبارات المتكافئة؟

٥- عنـــد فحــص زوج من الدرجات على نفــس المتغير لنفس الأفراد، تــم تطبيقه في وقتين مختلفين (على ســبيل المثال، درجة الاختبار القبلي تســمى X1 ودرجة الاختبار البعدي تســمى X2)، كيف يمكن لقوة الارتباط بــين الدرجات (أي الارتباط الذاتي الخاص بهم) أن تؤثر على ثبات درجة التغيير المحســوبة بطرح درجة من الأخرى. على ســبيل المثال، (X2– X1)؟

٦- تمثل جميع معاملات الثبات تقريباً نســبة من نوع واحد من المعلومات بالنســبة لنوع آخر. في الواقع، هذه النســبة تشــكل أساسًــا تعريف الثبات. اشرح باختصار ما هما نوعا المعلومات الأساسيَّان لتعريف الثبات وأنه، عند التعبير عنهما كنسبة، هما الأساس لمعاملات الثبات.

٧- ما هي بعض الانتقادات لاستخدام معامل ألفا؟ ومتى تنطبق هذه الانتقادات؟ وما هي العلاجات أو البدائل الموجودة للاستخدام التقليدي للمعامل ألفا، ومتى تكون مناسبة؟

ملاحظات:

١- بالنســبة للبنــود الموزونة، يتم ضرب التغايرات في حاصــل ضرب التباينات في مربعات أوزان البنــود المقابلة لهــا. انظــر (Nunnally, 1978, pp. 154-156) للحصول على وصف أكثر وضوحاً.

٢- خلال الكتاب، سوف يظهر حل أي تمرين يتطلب إجابة رقمية في حاشية.

٣- الإجابــات: (أ) ١,٢، ١,٠، و١,٨ (التــي تبلــغ ٤,٠)؛ (ب) ٧,٠ (مجمــوع جميع البنود في المصفوفة)؛ (ج) (٣/٢) * [١-(٤,٠ /٧,٠)] = ٠,٦٤.

الفصل الرابع
الصدق

في حـين يتعلق الثبات بمدى تأثير متغير على مجموعة من البنود، فإن الصدق يتعلق بما إذا كان المتغير هو السبب الأساسي لتغاير البنود. وبقدر ما يكون المقياس ثابتاً، فإن التباين في درجـات المقياس يمكن أن يُعزى إلى الدرجة الحقيقية لبعـض الظواهر التي تمارس تأثيراً سببياً على جميع البنود أو الفقرات. ومع ذلك، فإن تحديد درجة ثبات المقياس لا يضمن أن المتغير الكامن الذي تتشارك البنود فيه، في الواقع، هو المتغير المهم لمطور المقياس. وأن مدى كفاية المقياس كمقياس لمتغير محدد (مثل الضغوط النفسية المدركة) هي مسألة صدق.

وقـد أعطى بعض المؤلفين معنى أوسـع للصدق. على سـبيل المثال، وصف ميسـيك Messick (1995) سـتة أنـواع أو أنمـاط مـن الصـدق، أحدهـا (الصـدق التتابعـي consequential validity) يتعلـق بتأثـير الطريقة التي يتم بها اسـتخدام درجاتهم على المجيبـين. وعلى الرغم من أن آراء "ميسـيك" بشـأن الصدق أثارت بعـض القضايا المثيرة للتفكير، فإن نظام التصنيف الخاص به لم يُعتمد على نطاق واسـع. ووفقاً للتفسـير الأكثر تقليدية (المحافظ)، يتم اسـتنتاج الصدق من الطريقة التي تم بها بناء وتكوين المقياس، أو قدرتـه على التنبؤ بأحداث محددة، أو علاقته بمقاييس مفاهيم بنائية أخرى. وهناك ثلاثة أنواع أساسية من الصدق التي تتوافق مع هذه العمليات وهي:

١- صدق المحتوى Content validity.

٢- الصدق المرتبط بمحك أو بمعيار Criterion-related validity.

٣- الصدق البنائي أو صدق التكوين الفرضي Construct validity.

وسـيتم اسـتعراض كل نـوع بإيجـاز. وللحصول عـلى معالجة أكثر شـمولاً للصدق، بمـا في ذلك مناقشـة القضايـا المنهجيـة والإحصائيـة المتعلقـة بالصـدق المرتبط بمحك ومـؤشرات الصـدق البديلة، انظـر الفصل العاشر في كتاب "غيسـيلي، وكامبـل، وزيدك" Ghiselli, Campbell, and Zedeck (1981). وقـد يرغب القراء أيضاً في النظر في عرض ميسيك (١٩٩٥) الأكثر شمولاً لموضوع الصدق.

صدق المحتوى أو المضمون:

يتعلق صدق المحتوى بكفاية عينة البنود - أي المدى التي تعكس فيها مجموعة محددة من البنود نطاق أو مجال المحتوى. ويكون تقييم صدق المحتوى أسهل عندما يكون المجال أو النطـاق (على سـبيل المثال، جميع مفردات الكلمات التي يتم تدريسـها لطلاب الصف السـادس) محدداً بشكل جيد. وتكون المسألة أو المشكلة أكثر صعوبة عند قياس السمات مثـل المعتقدات أو المواقف والاتجاهـات أو التصرفات لأنه من الصعـب تحديد بالضبط ما هـو نطـاق البنود المحتملة ومتى تكون عينة من البنود ممثلـة للنطاق أو المجال. من الناحيـة النظريـة، يتمتع المقياس بصدق المحتوى عندما تكـون بنوده مجموعة فرعية تم اختيارها عشوائيًا من مجموعة كلية من البنود المناسبة. في مثال اختبار المفردات المستخدم أعـلاه (مفردات الكلمات التي يتم تدريسـها لطلاب الصف السـادس)، يمكن تحقيق ذلك بسـهولة. حيث سـيتم تعريف كل الكلمات التي تم تدريسـها خلال العام الدراسي على أنهـا المجموعة الكلية للبنـود. وبعد ذلك يمكن أخذ مجموعة فرعيـة من هذه الكلمات كعينة. ومع ذلك، في حالة قياس المعتقدات، على سـبيل المثال، لا توجد لدينا قائمة ملائمة من البنود ذات الصلة. إلا أن أحد أسـاليب أو طرق تطوير مقياس ما (على سـبيل المثال، مراجعـة البنـود من قبل الخبراء من حيث صلتها بمجال الاهتمام، كما هو مقترح في الفصل الخامس) يمكن أن تساعد في زيادة ملاءمة البند. على سبيل المثال، إذا كان الباحث بحاجة إلى وضـع مقياس يتناقض مـع النتائج المتوقعة والنتائج المرجوة (على سـبيل المثال، توقع إشراك المريـض في صنع القرار مقابل رغبة الطبيـب في إشراك المريض في صنع القرار)، فقد يكون من المستحسـن إثبات أن جميع النتائج ذات الصلة كانت ممثلة في البنود. وللقيام بذلك، قد يكون للباحث زملاء على دراية بسـياق البحث يقومون بمراجعة قائمة أولية من البنود واقتراح مجـالات المحتوى التي تم إهمالها ولكن ينبغي إدراجها في المقياس. ويمكن بعد ذلك إضافة بنود تعكس هذا المحتوى.

يرتبط صدق المحتوى ارتباطًا وثيقًا بتعريف المفهوم البنائي قيد الدراسة. وبكل بساطة، يجـب أن يعكـس محتوى المقيـاس التعريف المفاهيمـي أو التصـوري المنطبق على هذا المقيـاس. قد تكـون بعض المفاهيم قد تـم تعريفها بأكثر من طريقة مـن قبل المنظرين أو قد تقع عند تقاطع أو التقاء مفاهيم متعددة. ومن الضروري أن يتضمن محتوى البنود

جوانب الظاهــرة التي تم توضيحها في تعريفها المفاهيمي أو التصوري وليســت الجوانب الأخرى التي قد تكون ذات صلة ولكنها خارجة عما يقصده الباحث لتلك الأداة المعينة.

وكمثــال على ذلك، قام «ســتربا وزمــلاؤه» Sterba and colleagues (2007) بتطوير مقيــاس الفعالية الزوجية المتعلقــة بكيفية إدراك الأزواج الذين أصيــب أحدهم بالتهاب المفاصــل الروماتويدي لقدرتهــم على إدارة المرض كفريق واحــد. وكانت هذه الأداة هي الأولى لتقييــم الفعالية الزوجية، وثقة الزوجين في قدرتهم كفريق واحد على اتخاذ مختلف الإجراءات لتعزيز الصحة. وهكذا، كان المفهوم الأساسي للأداة مختلفًا عن المفاهيم الأخرى مثــل الفعاليــة الذاتية الفردية، على الرغم من أنه ذو صلــه بها. وكانت المرحلة الأولى من هذا الجهد عبارة عن دراســة لتطوير البنود والتي شــكلت الأساس لادعاءات المؤلفين حول صدق المحتوى. وتهدف هذه الدراسة إلى تحديد المحتوى المناسب من الأدبيات التجريبية والنظريــة الأوســع نطاقًا لإدراجها في المقيــاس. على الرغم من أن المؤلفــين قاموا بفحص المحتوى لمقاييس المفاهيم البنائية ذات الصلة (على ســبيل المثال، الفعالية الذاتية الخاصة بالتهــاب المفاصــل)، فقد قاموا بتوجيــه تطوير البنود الخاص بهم إلى ســمات أو خصائص محددة للمفهوم البنائي كما حددوه. وتنطوي أحد الجوانب المهمة من هذا التعريف على مدى شــعور الأزواج بالثقة بأنهم، كفريــق واحد، يمكنهم إدارة تحديات المرض. وبناءً على ذلك، فقد اشــتملت دراســة تطوير البنود على مقابلات مع الأزواج لمعرفة أفكارهم حول مــا إذا كان المفهوم البنائي يتوافــق معهم، ولفهم كيفية تصورهم لــه، وتحديد اللغة التي اســتخدموها لوصفه. إن الأفكار والرؤى المســتقاة من التعريــف النظري للمفهوم البنائي، ومراجعــة الأدبيات، والمقابلات مع المرضى تم الاسترشــاد بها عند بناء وتكوين البنود. وقد تمت كتابة البنود بشكل صريح للتعبير عن جانب تصور فريق الأزواج فيما يتعلق بالفعالية. وقــد ضمنــت هذه العمليــة أن محتوى البنود يعكــس المفهوم البنائي المحــدد الذي كان الباحثــون مهتمين به وليس مفاهيم أخرى مختلفة (مثل الفعالية الذاتية أو الدعم الزوجي المــدرك) والتي قد تكون ذات صلة بها من الناحية النظرية. وكخطوة إضافية للتحقق من صدق المحتوى، فقد طلب «ستربا وآخرون»، Sterba et al. (2007) من مجموعة من خبراء المحتــوى مراجعة البنود التــي طورها فريق البحث في البداية في ضــوء تعريفهم النظري. وكان هــذا الإجــراء بمثابة تحقق إضــافي من أن البنود كانت تمثل المحتــوى ذا الصلة الذي

تم تصميم الأداة لقياســه. وإجمالاً، فقد زادت هذه الخطوات من احتمال تضمين وإدراج المحتــوى ذي الصلــة في المقياس في حــين لم يتم إدراج المحتوى غــير ذي الصلة، مما يدعم الادعاءات بصدق المحتوى.

نطاق المتغير والآثار المترتبة على صدق المحتوى:

ثمة مسألة أو قضية مرتبطة ارتباطًا وثيقًا بصدق المحتوى تتمثل في نطاق scope المتغير موضــع الاهتمام وبالتالي البنــود التي تؤلف المقياس. من المحتمل أن يكون الباحث مهتمًا بمتغير شــامل مثــل نوعية الحياة أو تفضيــلات الشراء أو بمتغير أكثر تحديــدًا (ضيقًا) مثل السـعادة أو الوعي بالتكلفة. وبطبيعة الحال، فإن المحدد الأســاسي لما نقيســه ينبغي أن تكون الأســئلة التي نسعى للإجابة عليها وفهمنا للقضايا النظرية ذات الصلة بتلك الأسئلة. وفي الوقت نفســه، ينبغي أن نضع في اعتبارنا مدى صلة اتســاع المفهوم البنائي الخاص بنا بصدق المحتوى.

ومــن الواضح، فإن المحتــوى المبالغ فيه أكثر من اللازم أو المحتــوى المختصر الأقل من اللازم يمكن أن يثير مشاكل إذا أدى إلى وجود عدم تطابق بين المفهوم الذي يراد من المقياس قياســه ونطاق البنود التي يحتوي عليها. في ســياق الاختبارات التعليمية، حذر "ميسيك" Messick (1995) من المخاطر المحتملة لـ "ســوء أو نقص تمثيل المفهوم"، أي أخذ عينات محــدودة جداً من مجال ونطاق المحتــوى، وبالتالي من المحتمــل أن يحرم أحد الممتحنين من فرصة إثبات وإظهار معرفته أو معرفتها. ومع ذلك، يشــير أيضاً إلى خطر "التباين غير الملائم للمفهوم البنائي" الذي يمكن أن ينتج عن أخذ عينات من المحتوى على نطاق واســع جداً. على الرغم من أن "ميسيك" يناقش هاتين الظاهرتين على أنهما تهديد للصدق البنائي، إلا أنهما تتعلقان بوضوح باتساع نطاق اختيار المحتوى وبالتالي صدق المحتوى.

وعلى الرغم من أن إســتراتيجية اختيار عينة المحتوى والتي تكون إما مفرطة في الشمول أو شــديدة التقييد يمكن أن تمثل إشــكالية، ويبدو أن هناك تفاوتاً أو عدم تناسق، مع وجود مفاهيم بنائية أوسع نطاقاً ومجموعات بنود أكثر شمولاً. ومن مظاهر هذا الاتجاه أنه يبدو، في بعض السياقات البحثية، أن هناك تفضيلاً متزايداً لمقاييس القياس ذات الأهداف العريضة أو التــي تكون على نطاق واســع. والقياس المتعلــق بالحالة الصحية هــو مثال على ذلك.

ومــن الواضح أن هناك بعض الأســباب التي تقف وراء هذا التفضيــل المتزايد للمقاييس الأكثر شــمولاً. إذا كان المقياس يمكن أن يغطي سمة واسعة نسبياً، فمن المحتمل أن يكون أكثر فائدة واستخداماً. ويمكن لشخص أن يقول، على سبيل المثال، بأن مقياساً عاماً للإعاقة قــد يكون أكثر جاذبية من مقياس خاص بمجموعة معينــة من المرضى. وهذا لا يغني عن الحاجــة إلى وضــع وتطوير مقياس جديد لكل نوع من الأمراض المختلفة فحســب، بل إنه يعزز المقارنات بين المجموعات لأنه يمكن قياس كليهما باستخدام نفس الأداة. وعلاوة على ذلــك، فإن أدوات القياس الأكثر قابلية للتطبيق يمكــن أن تؤدي مهمة أو وظيفة مزدوجة، وهــي تلبيــة الاحتياجات الإدارية فضــلاً عن الاحتياجــات البحثية. ولأغــراض مثل تلبية ومعالجة حقوق ومطالبات الإعاقة، على سبيل المثال، يمكن تطبيق أداة أكثر عمومية بشكل موحد لقياس الإعاقة عبر الظروف (الحالات) ومن ثم تحديد الأهلية للحصول على تعويض بطريقــة موحدة بين المرضى والســياقات (الحالات). كذلك، قد تســتفيد مقارنات فعالية العلاج عندما يتم قياس نتائج العلاج بنفس الأداة. ومع ذلك، قد تخفي هذه المزايا الظاهرة أهمية الاختلافات السياقية والعينات المحددة في طريقة أداء المقياس.

وكما هو الحال بالنسبة للثبات، فإن الصدق ليس خاصية متأصلة في أداة القياس بل هو خاصية لأداة القياس في ســياق اســتخدامها. قد تكون الأداة صادقة في سياق واحد ولكنها غير صادقة في سياق آخر أو عند استخدامها بشكل مختلف. كمثال واضح للغاية أو بديهي على ذلك، فإن جهاز ضغط الدم في حالة عمل جيدة هو وســيلة صادقة لتقييم الضغط في الأوعية الدموية للشــخص؛ ومع ذلك، إذا تم وضع الجهاز حول رأس شــخص ما في محاولة لقياس ذكائه، فإن النتيجة لن تكون صحيحة أو صادقة. لا يوجد الصدق في الأداة نفســها ولكــن في كيفية وطريقة اســتخدامها. وينطبق هذا أيضًا عــلى المقاييس في القياس القائم على البنود. ولذلك، فإن صدق محتوى أداة ما سيعتمد أو يتوقف على مدى جودة وسلامة البنود التي تتألف منها خريطة المقياس للمفهوم البنائي موضع الاهتمام للمجتمع الإحصائي وسياق البحث المحدد. وسيؤثر تنوع محتوى البند على خريطة العمل هذه.

يتمثــل أحد أوجه القصــور المحتملة في الأداة التي تحاول قيــاس مفهوم ما على نطاق واسع في أن البنود التي تم اختيارها لتمثيل وتجسيد هذا المفهوم قد لا تنطبق بالتساوي في جميع المواقف والحالات أو على كل المجتمع الإحصائي. في الواقع، قد يكون صدق المحتوى

للأداة معتمداً إلى حد كبير على السـياق. لنفكر، على سبيل المثال، أداة افتراضية تهدف إلى تقييـم (قياس) الوعي بالتكاليف. ولأغـراض تدريبية أو تعليمية، دعونا نفترض أن الأعمال السـابقة قد كشفت عن أن بعض الأفراد مدركون بدرجة عالية للتكلفة عند إجراء عمليات الشراء. سـوف يمتنع هؤلاء الأفراد عن إجراء عملية الشراء، حتى لو كانت السلع المعروضة جذابـة للغاية، إذا كانوا يعتقدون أن التكلفة لا تمثل قيمة اسـتثنائية. ولنفترض كذلك أن اشخاصاً آخرين كان وعيهم أقل بمتغير التكلفة. بالنسبة لهؤلاء الأفراد، فإن التكلفة ليست عاملاً محدداً رئيسـياً في قراراتهم المتعلقة بالشراء، ولكنها ليسـت سـوى واحدة من عدة اعتبارات مهمة تشـمل الجودة والرغبة. وإذا كانت سـلعة معينة من السـلع الاستهلاكية جذابة للغاية بالنسبة لهم، فقد يفضلون الحصول عليها عندما تتاح الفرصة نفسها بدلاً من انتظار إمكانية العثور على سعر أقل لنفس السلعة أو المماثلة لها في أماكن أخرى.

ومع الاسـتمرار في هذا المثال الافتراضي، دعونا نفـترض أن الباحث يريد تكوين مقياس لتقييـم الوعـي بالتكاليف على النحو المبين هنا. قد يحقق ذلـك من خلال التفكير بعناية حول ما يشمله المفهوم ثم تطوير البنود التي تعكس ذلك التعريف العملي للمفهوم.

ومن المسـائل أو القضايا التي سـيواجهها الباحث مدى اتسـاع أو ضيق نطاق قرارات الشراء التي ينبغي أن تتضمنها وتشـملها الأداة. وهذا، بدوره، ينبغي أن يتم الاسترشـاد به في تكوين وبناء البنود. وفي إحدى الحالات المتطرفة، يمكن تركيز البنود بشـكل محكم إلى حد ما على خيارات شراء محددة، مثل شراء سـيارة جديدة. وعلى النقيض من ذلك، يمكن للباحث أن يحاول وضع مقياس عام يمكن تطبيقه على أي نوع من الحالات التي قد تكون فيها التكلفة قضية أو مشكلة، بما في ذلك مجموعة كبيرة من السلع والخدمات. دعونا ننظر في بعض الأمثلة على أنواع البنود التي قد يسفر عنها كل نهج من هذين النهجين. افترض أن مجموعتي البنود التاليتين طلبتا من المجيبين الموافقة على كل بند على مقياس اتفاق مكون من ست نقاط مع خيارات استجابة تتراوح بين لا أوافق بشدة إلى أوافق بشدة. البنود أرقام ١ و٢ و٣ التالية تخص الوعي بالتكلفة في سياق شراء سيارة.

١- أفضل تأجيل الحصول على السيارة التي أريدها حقاً بدلاً من دفع بنس واحد زيادة لذلك.

٢- أفضل أن أدفع أكثر قليلاً وأحصل على السـيارة التي أريدها بالضبط بدلاً من توفير بعض المال ولكن يجب المساومة.

٣- أنا أعتبر السعر الأول الذي يذكره مندوب مبيعات السيارات ليس أكثر من حيلة مساومة.

وتتضمن المجموعة الثانية من البنود (أي من البند رقم ٤ إلى البند رقم ٦) نسخة عامة من المجموعة الأولى، وصيغت لتكون قابلة للتطبيق عبر السياقات المختلفة.

٤- أنــا أفضل تأجيل امتــلاك ما أريده حقاً بدلاً من دفع بنــس واحد أكثر لذلك مما يجب عليّ دفعه.

٥- أفضل أن أدفع أكثر قليلاً وأحصل على السلع أو الخدمات التي أريدها بالضبط بدلاً من توفير بعض المال ولكن يجب المساومة.

٦- أنا أعتبر أن السعر الأول المذكور لأي سلع أو خدمات ليس أكثر من حيلة مساومة.

ويبدو أن البنود الثلاثة الأولى تستفيد من الوعي بالتكلفة وتتشارك في وجهة نظر معينة حول أســعار الســيارات وأهميتها في اتخاذ خيارات شراء السيارات. قد يوافق بعض الناس بشدة أو لا يوافقون على العبارات بصيغتها المكتوبة، ولكن ليس من الصعب أن نتصور أن الاســتجابات التي اختارها شخص ما ستكون متشــابهة إلى حد ما عبر مختلف حالات شراء الســيارات. كما يبدو مــن المعقول أن تنطبق هذه البنود بشــكل أو بآخر بنفس الطريقة تقريباً على غالبية الأشــخاص الذين يواجهون قرار شراء ســيارة. وعلى الرغم من أن مدى تأييد مختلف الناس للفقرات من شأنه أن يختلف على الأرجح اختلافاً كبيراً، إلا أنني أتوقع درجة عالية نسبياً من الاتساق في مستويات الموافقة التي يقدمها أي شخص.

دعونا الآن ننتقل إلى العبارات الثلاث المتبقية، أي البنود من رقم ٤ إلى رقم ٦. وليس من الصعب تصور أن مجال المتغير الأوســع نطاقاً (أي النطاق الشــامل، وليس المحدد بالسياق للوعــي بالتكلفة) وما يرتبط به من بنود قد يســبب بعض الصعوبات. ولا أتوقع أن يتفق معظم الناس على هذه البنود على نحو مماثل عبر ســياقات مختلفة بشــكل ملحوظ. على سبيل المثال، ينبغي النظر في كيفية استجابة الناس على هذه البنود في ثلاث حالات أو مواقف مختلفة: (أ) شراء تلفزيون، (ب) شراء عمل أو فن نادر، (ج) تقرير ما إذا كان ينبغي تلقي المشــورة القانونية أو التخلي عنها في مســألة ذات أهمية حيوية. وإنني أشك بشدة في أن هذه الســياقات المختلفة ستحســن علاقة هذه البنود مع متغيرهــا الكامن المقصود منها، وأن معرفة كيفية اســتجابة شــخص ما في أحد تلك الســياقات لن تكون بالضرورة تنبؤيه

لكيفية اســتجابته للســياقين الآخرين. وأساس هذا الشك هو اعتقادي بأن الوعي بالتكلفة كما تعكســه هذه البنود الأخيرة له معنــى مختلف اختلافاً جوهرياً في ســياقات مختلفة. وأهميته بالنســبة للمتغيرات الأخرى التي تؤثر على الاســتجابات للبنود تكون غير متسقة. ويبدو أن الوعي بالتكلفة الذي يحفز المســاومة (التفاوض) على تكلفة المشــورة القانونية التي قد تؤدي إلى بعض العواقب الخطيرة مثل السجن أو الخسارة الكاملة لأصول المرء هو متغير مختلف تماماً عن المتغير الذي ينطوي عليه التفاوض والمســاومة حول سعر تلفزيون جديد. وإذا أكدت الأدلة التجريبية هذه الشــكوك، فإنني سأقول إن المجموعة الثانية من البنــود الأكثر عمومية لم تكن ذات مضمون صحيح عبر الســياقات. وعلاوة على ذلك، أود أن أقترح أن عدم صدق الأدلة في ســياقات معينة (مثل تحديد الخيارات المتعلقة بالحصول على مســاعدة قانونية أو طبية هامة) ينشــأ، على الأقل جزئياً، عن عدم خصوصية وتحديد المفهوم البنائي. ونتيجة لهذا النقص في التحديد، قد يتعرض صدق محتوى البنود من البند رقم ٤ إلى البند رقم ٦ للخطر.

وبطبيعــة الحال، يمكن للمــرء أن يجادل بأن البنــود من البنــد الأول إلى البند الثالث ستكون أيضًا محتوى غير صالح في مواقف أو حالات أخرى غير شراء سيارة. وعلى الرغم من أن هذا صحيح بلا شك، إلا أن الفرق المهم هو أن القيد في صدق محتوى تلك البنود واضح، وأن الأشخاص العقلانيين لن يحاولوا استخدامها لتقييم الوعي بالتكلفة في سياقات مختلفة بشــكل ملحوظ. ومع ذلك، في حالة وجود البنود الأكثر عمومية، فإن عدم ملاءمتها لبعض المواقف يكون أقل وضوحًا ويمكن بسهولة ألا يلاحظه أحد.

في المثــال الســابق، كان صدق المحتوى معرضاً للخطر عندما تم اســتخدام نســخ أكثر عموميــة مــن البنود لأن مفهوم الوعي بالتكلفة يأخذ في الأســاس معنــى مختلفًا في بعض الســياقات. وثمة طريقة أخرى يتم من خلالها تعريف أوســع نطاقــاً للمفهوم البنائي ربما يعــرض صــدق المحتوى للخطر وهي عندمــا تكون مجموعة فرعية مــن البنود ذات صلة بالمفهوم البنائي أحيانًا وأحيانًا لا تكون كذلك. وهذا يعني، أنه في بعض الظروف، ســتكون بعض البنود مصدرًا لحدوث تباين غير ذي صلة بالمفهوم البنائي (ميسك، Messick, 1995)، حتــى عندمــا تحتفظ بنود أخرى بصلتها وأهميتها. وكمثال افــتراضي آخر، ضع في اعتبارك مجموعة فرعية من البنود قيد النظر لمقياس عام للإعاقة. (مرة أخرى، يمكننا أن نفترض أن

خيارات الاستجابة المصاحبة تتراوح من لا أوافق بشدة إلى أوافق بشدة).

أ - لدي مشكلة في استخدام يدي للمهام التي تتطلب قوة قبضة.

ب- لدي مشكلة في الوقوف من وضعية الجلوس.

ج - عندما أبقى في وضع واحد لفترة طويلة جداً، فإن أطرافي تصبح صلبة متيبسة.

د - أنا أتعب بسهولة عندما أبذل جهداً كبيراً جداً.

هـ - لم يعد بإمكاني القيام بأشياء كان بإمكاني القيام بها بسهولة.

يمثل كل بند من هذه البنود قيداً وظيفياً يمكن أن يعتبر ممثلاً للإعاقة بصورة منطقية. وبالنسبة لبعض أنواع المرض، قد يتوقع أن تكون كلها مظاهر للحالة وقد تميل إلى أن تحدث في انسجام. غير أن هناك مشكلة تنشأ إذا كان الغرض من البنود هو تقييم الإعاقة بالنسبة لحالة لا تتضمن كل هذه القيود. وفي مثل هذه الحالة، قد تكون بعض البنود تتمتع بصدق المحتوى في حين أن البعض الآخر ليس كذلك. لذلك، على سبيل المثال، بالنسبة لشخص يعاني من كسر حاد في الساق، فإن البند رقم «أ» والذي يصف عدم وجود قوة القبضة سيكون بنداً لا يتمتع بصدق المحتوى. أي إن قوة القبضة ستكون غير مفيدة فيما يتعلق بتأثر كسر ساق الشخص. في المقابل، بالنسبة لبعض الحالات الأخرى، ربما السكتة الدماغية أو التهاب المفاصل الروماتويدي، فإن بنداً عن قوة القبضة قد يتمتع بصدق المحتوى. لذلك، في حين أن إدراج وتضمين البند رقم «أ» يضر بصدق المحتوى لمن يعاني من ساقة المكسورة، فإن حذفه أو إهماله قد يضعف صدق المحتوى للمريض الذي يعاني من التهاب المفاصل الروماتويدي. وتكمن الصعوبة هنا في أن التعريف الإجرائي الضمني للإعاقة الذي اعتمدته عند تكوين هذه البنود التوضيحية كان فضفاضاً وواسعاً أكثر مما ينبغي. ونتيجة لذلك، وضعت قائمة بمظاهر الإعاقة المقبولة التي قد تظهر في مجموعة من الحالات الحادة أو المزمنة دون مراعاة للكيفية التي يمكن أن تناسب هذه القائمة، أي حالة محددة تؤدي إلى الإعاقة. ومن ثم، فإن التعريف الأوسع نطاقاً يؤدي مرة أخرى إلى نشوء حالة يمكن أن تضر بصدق المحتوى.

تم وضع الأمثلة السابقة لتوضيح وإثارة نقطة. وإنني أشك في أن أي شخص سيفكر فعلياً في وضع وتطوير مقياس للوعي بالتكلفة مع فكرة أنه يمكن استخدام بنود مثل البنود من

البند الرابع إلى البند السادس في أي سياق دون عقاب (مخاطرة). كما أن أي باحث خبير (ذي خبرة) لن يجمع بين البنود من البند الأول إلى البند الخامس في مقياس واحد يهدف إلى تقييم الإعاقة عبر مجموعة واسعة من الظروف والحالات. ومع ذلك، توضح هذه الأمثلة المتطرفة نقطة تنطبق في بعض حالات القياس في العالم الحقيقي أو من واقع الحياة: عندما يكون المتغير والمقياس المراد استخدامه ذو نطاق واسع للغاية، يزداد احتمال عدم تطبيق بعض المؤشرات (وبالتالي لن يكون له صدق محتوى) في حالات معينة أو لمجموعات معينة.

ومن النتائج السلبية الأخرى المحتملة لتعريف متغير ما على نطاق واسع أن خطوات إثبات صدق المحتوى قد يصبح أكثر عبئاً وإرهاقاً. فعلى سبيل المثال، كما هو موضح في الفصل الخامس، من الشائع أن يطلب من الخبراء مراجعة محتوى البنود كوسيلة لضمان صدق المحتوى. ومع اتساع نطاق المفهوم البنائي، قد تحتاج أيضاً الخبرة الجماعية للأشخاص الذين طُلب منهم مراجعة صدق المحتوى إلى التوسع. كما أن المهمة المعروضة على هؤلاء الخبراء قد تصبح أكثر صعوبة لأن هؤلاء المحكمين يضطرون إلى النظر في متى قد يكون بند معين ذا صلة بالمفهوم البنائي أو لا يكون له صله. وهذا يتطلب أن يتصوروا أو يتخيلوا شيئاً ما يقترب من جميع الحالات والمواقف الممكنة التي يمكن فيها استخدام الأداة وتقييم مدى ملاءمته لكل منهما - وهو مهمة وعمل هائل. ومع تناقص خصوصية المفهوم البنائي بدقة، قد يجد المحكمون صعوبة أكبر في تقييم ما إذا كانت هناك بنود محددة تناسب المفهوم البنائي على نحو كافٍ. وقد يتطلب ذلك إنشاء فريق من المحكمين الخبراء أكبر مما يلزم مع تحديد أكثر دقة للمفهوم البنائي لتحقيق نفس الدرجة من الثقة في تقديراتهم وتقييماتهم.

وعندما تتعلق أسئلة البحث ذات الاهتمام بمتغيرات واسعة النطاق بطبيعتها، قد يفكر الباحثون فيما إذا كان مقياس عام واحد أو مقاييس محددة متعددة تكون أكثر ملاءمة للمهمة وأكثر ملاءمة لإثبات صدق المحتوى. ومواصلةً لمثال الوعي بالتكاليف، قد يفكر الباحث فيما إذا كانت عدة مقاييس خاصة بسياقاتها الفريدة من نوعها ستكون بديلاً أكثر فائدة لمجموعة غير محددة من البنود المماثلة لمجموعة البنود من البند الرابع إلى البند السادس السابق ذكرها. أو قد يكون من المفيد، في حالة مقياس الإعاقة، اختيار مجموعات

البنود التي تشـير تحديداً إلى أنواع القيود التي يمكن توقعها بين مجموعة محددة بشـكل ضيق للغاية من الناس.

يختلف صدق المحتوى عن الأنواع الأخرى من الصدق التي سنناقشـها لأنه يُعرّف فعلياً بالإجـراءات التي يتخذها منشـئ أو مبتكر المقياس في بدايـة تطوير المقياس. وأهم هذه العوامل هو النظر بعناية في المتغير الذي يقصد من المقياس أن يقيسه ومن ثم أنواع البنود التي تكون عينات كافية من المجال أو النطاق الذي يتضمنه المتغير. وتسـتحق المفاضلات بـين النطاقات الأكثر عمومية والنطاقات الأكثر تحديداً ومجموعات البنود اهتماماً دقيقاً في هـذه المرحلة من عملية تطوير المقيـاس ويمكن أن يكون لها تأثير كبير على صدق محتوى البنـود التـي تم اختيارها في نهايـة المطاف. وبالمثل، ينبغي للباحـث، عند اختيار مقياس موجود بالفعل لاسـتخدامه في دراسـة ما، بـدلاً من مجرد الاعتماد على التقارير السـابقة المتعلقة بصدق المحتوى من الدراسـات السابقة، أن يفكر بعناية في كيفية توافق أو تطابق محتوى المقاييس المتاحة مع المجتمع الإحصائي، والسـياق، وفحص أسئلة الدراسة المحددة التي تم التخطيط لها.

ونظراً لأن اتسـاع المتغير الذي يختار الباحث دراسـته يحدد بطريقة مباشرة وشـفافة حجم المجموعة المحتملة من المحتوى الذي يجب أن يمثلها مطور المقياس بشـكل مناسب، فقد فضلت مناقشة هذه القضية جنباً إلى جنب مع صدق المحتوى. ومع ذلك، فإن اتساع النطاق الذي يأمل مقياس ما أن يشـمله، تترتب عليه أيضاً آثار للأنواع الأخرى من الصدق التي أناقشها في الجزء المتبقي من هذا الفصل. ولن أكرر ما قمت بتغطيته حتى الآن، ولكن يجب أن يدرك القراء أن ملاءمة ومناسـبة البنود للمفهوم البنائي أمر بالغ الأهمية بالنسبة لجميع أنواع الصدق التي تم اسـتعراضها في هذا الفصل. ومع اتسـاع حدود هذا المفهوم البنـائي، يمكن أن تصبح مهمة إيجاد وإثبات الصـدق المرتبط بالمحك (بالمعايير) أو الصدق البنائي أكثر صعوبة، تماماً كما هو الحال بالنسبة لصدق المحتوى.

الصدق المرتبط بمحك (معيار):

من أجل الحصول على الصدق المرتبط بمحك أو بمعيار، كما يوحي المصطلح، لا يشــترط إلا أن يكــون لبند أو مقياس ارتباط تجريبي ببعض المحكات أو "المعيار الذهبي" المفترض. وســواء قد تم فهم الأســاس النظري لذلك الارتباط أم لا، فإن ذلــك ليس له صله بالصدق المرتبط بالمعيار. على سبيل المثال، إذا تبين أن عملية التغطيس (البحث بعصا الاستنباء عن الماء) مرتبطة بشــكل تجريبي بتحديد موقع مصادر الميــاه الجوفية، فإن عملية التغطيس ســتكون صادقة فيما يتعلق بمعيار حفر الآبار بنجاح. وبالتالي، فإن الصدق المرتبط بالمحك في حد ذاته هو مســألة عملية (تطبيقية) أكثر من كونها مســألة علمية، لأنها ليست معنية بفهــم العملية بل مجرد التنبؤ بها وتوقعها. وفي الواقع، غالباً ما يشــار إلى الصدق المرتبط بمحك على أنه صدق تنبؤي predictive validity.

لا يعني الصدق المرتبط بمعيار تحت أي اسم بالضرورة وجود علاقة سببية بين المتغيرات، حتــى عندما يكون الترتيب الزمني للتنبــؤ والمعيار واضح لا لبس فيه. وبطبيعة الحال، قد يكون التنبؤ في سياق النظرية (مثل التنبؤ كفرضية) ذا صلة بالعلاقات السببية بين المتغيرات ويمكن أن يخدم أو يؤدي غرضًا علميًا مفيدًا.

وثمــة نقطة أخــرى جديرة بالملاحظة بشــأن الصدق المرتبط بمعيار، هــي أن المرء، من الناحية المنطقية، يتناول نفس نوع قضية الصدق سواء كان المعيار يتبع أو يسبق أو يتزامن مــع القياس المعني. وهكذا، بالإضافة إلى الصدق التنبؤي، فإن الصدق المتزامن أو التلازمي concurrent validity (على ســبيل المثال، التنبؤ بمهارة القيادة من الإجابات على الأســئلة الشــفوية التي طرحت في اختبار القيادة) أو حتــى الصدق اللاحق postdictive validity (على ســبيل المثال، التنبؤ بالوزن عند الولادة من مقياس حالة النمو للطفل الرضيع) يمكن اســتخدامها بشــكل مترادف إلى حد ما مع الصدق المرتبط بمحــك أو معيار. وأهم جانب مــن جوانب الصدق المرتبط بمحك أو معيار ليس العلاقة الزمنية بين المقياس المعني (الذي يراد التثبت من صدقه) والمعيار الذي يحاول المرء استنتاجه من قيمته، بل هو قوة العلاقة التجريبيــة بين الحدثين. ويتمتع الصدق المرتبــط بمحك بميزة مقارنة بالمصطلحات الأخرى تتمثل في كونه محايداً زمنياً، وبالتالي فهو الأفضل.

الصدق المرتبط بمحك مقابل الدقة:

وقبل الانتهاء من مناقشـــة الصـــدق المرتبط بمحك، هناك بضع كلــمات تتعلق بعلاقته بالدقة. وكما يشـــير «غيســـيلي وآخـــرون (1981) .Ghiselli et al، فـــإن معامل الارتباط، الـــذي كان المؤشر التقليدي للصدق المرتبط بمحك، قد لا يكـــون مفيداً للغاية عندما تكون الدقة التنبؤية هي القضية أو المســـألة. فعلى ســـبيل المثال، لا يكشف معامل الارتباط عن عـــدد الحالات التي يتم تصنيفها بشـــكل صحيح من قبل أحد المؤشرات (على الرغم من أن الجداول التي تقدم تقديرًا لنســـبة الحالات التي تقع في فئات مئوية مختلفة، اســـتنادًا إلى حجم الارتباط بين المنبئ والمعيار، تم وصفها من قبل"غيســـيلي وآخرون"، ص ٣١١). وقد يكـــون من الأنســـب في بعض الحالات تقســـيم كل من المتنبئ ومعيـــاره إلى فئات منفصلة وتقييـــم "معدل الدخول" لحالات الاســـتبدال إلى الفئة الصحيحة للمعيار اســـتناداً إلى فئة التنبؤ الخاصة بهم. فعلى سبيل المثال، يمكن تصنيف كل متغير إلى فئات «منخفضة» مقابل «عاليـــة» كـــما يمكن تصور الدقة على أنها نســـبة التصنيفات الصحيحـــة (أي الحالات التي تتطابق فيها قيمة المؤشر مع قيمة المعيار). حيث إن تقســـيم الشـــخص للفئات هو اعتبار مهم. لنأخذ في الاعتبار معياراً أو محكاً له حالتان غير تعســـفيتين، مثل «مريض» و«صحيح أو معـــافى»، وأداة تقييم تحتوي على مجموعة من الدرجات التي يريد المحقق تقســـيمها. والغرض من أداة التقييم هو التنبؤ بما إذا كان الأشـــخاص ســـيختبرون على أنهم إما موجب أو ســـالب بالنســـبة للمرض موضع الاهتمام. ونظرًا لأن النتيجة ســـتكون ثنائية الاستجابة، فمـــن المنطقي أن يكون المتنبئ ثنائي. هناك خطـــآن محتملان في التصنيف هما: يمكن أن يصنف المقياس عن طريق الخطأ شخصًا مريضًا حقاً على أنه صحيح (سليم) (نتيجة سلبية زائفة أو خاطئة) أو شخصًا صحيحاً حقًا على أنه مريض (نتيجة إيجابية زائفة أو خاطئة).

حيـــث يمكن أن يؤثر التقســـيم على معـــدلات هذين النوعين من الأخطـــاء على امتداد نطاق الدرجات من أداة التقييم. في النهايات، ســـيؤدي تصنيف كل شـــخص على أنه صحيح إلى تجنـــب أي ســـلبيات خاطئة أو زائفـــة (ولكن يزيد من الإيجابيـــات الزائفة) بينما يؤدي تصنيف كل شخص على أنه مريض إلى تجنب أي إيجابيات زائفة (ولكن يزيد من السلبيات الزائفـــة). ومن الواضح، في كلتا الحالتين المتطرفتـــين، أن أداة التقييم لن تكون لها أي قيمة

تنبؤيه على الإطلاق. والهدف، بالطبع، هو اختيار نقطة القطع التي ينتج عنها أقل عدد من الأخطــاء مــن أي نوع، وبالتالي أعلى درجة دقة. وفي كثــير من الأحيان، لا توجد نقطة قطع مثاليــة (أي، نقطــة قطع واحدة تؤدي إلى تصنيف مثالي). وفي مثــل هذه الحالة، قد يبذل الباحــث مجهودًا واعيــاً لتقليل نوع واحد من الأخطاء إلى الحد الأدنى بدلاً من الآخر. فعلى ســبيل المثال، إذا كان المــرض مدمراً وكان العلاج فعالًا وغير مكلــف وغير مضر، فإن تكلفة نتيجة سلبية زائفة (تؤدي إلى التفريط في العلاج) أكبر بكثير من تكلفة نتيجة إيجابية زائفة (تؤدي إلى الإفراط في العلاج). وبالتالي، فإن اختيار نقطة قطع لتقليل الســلبيات الزائفة مع قبول إيجابيات زائفة يبدو مناسباً. ومن ناحية أخرى، إذا كان العلاج مكلفًا ومزعجاً «كريهاً» في آن واحد وإذا كان المرض خفيفًا أو معتدلاً، فقد تكون المفاضلة العكسية أكثر منطقية.

أيضــاً، من المهم أن نتذكر أنه حتى لو كان الارتباط بين مقياس التنبؤ والمعيار تاماً، فإن النتيجة التي تم الحصول عليها في مقياس التنبؤ ليست تقديراً للمعيار أو المحك. معاملات الارتباط غير حساســة للتحولات الخطية لأحد المتغيرين أو كليهما. ويشير الارتباط المرتفع بــين متغيرين على أن الدرجات عــلى هذين المتغيرين والتي تم الحصــول عليها من نفس الشــخص سوف تشــغل مواقع مماثلة في توزيعات كل منهما. على سبيل المثال، شخص ما يســجل درجات عالية جداً على المتغــير الأول من المحتمل أيضاً أن يســجل درجات عالية جداً على المتغير الثاني إذا كان الاثنان مرتبطين بشــدة. ومع ذلك، فإن التعبير «عالية جداً» هــو مصطلح نســبي وليس مصطلحاً مطلقــاً ولا يأخذ في الاعتبار وحــدتي القياس الخاصة بالمتغيرين، على سبيل المثال. وقد يكون من الضروري تحويل وحدات قياس مقياس التنبؤ إلى وحدات القياس الخاصة بالمعيار للحصول على تنبؤ رقمي أو كمي دقيق. وهذا التعديل يعادل أو يساوي تحديد القاطع intercept المناسب بالإضافة إلى ميل خط الانحدار. ويمكن أن يــؤدي عــدم إدراك الحاجة إلى تحويل درجة ما إلى اســتنتاجات خاطئة. ومن المحتمل حدوث خطأ من هذا النوع على الأرجح إذا حدث أن تمت معايرة مقياس التنبؤ في وحدات تقع في نفس نطاق المعيار. افترض، على ســبيل المثال، أن شــخصاً ما وضع «مقياس مخالفة السرعة» التالي وذلك للتنبؤ بعدد تذاكر مخالفة السرعة التي سيحصل عليها السائقون على مدى خمس سنوات:

١- أنا أتجاوز الحد الأقصى للسرعة عندما أقود السيارة.

(كثيراً، أحياناً، نادراً، أبداً).

٢- على الطرق متعددة المسارات، وأنا أقود سيارتي أتجاوز الممر العابر.

(كثيراً، أحياناً، نادراً، أبداً).

٣- أنا أحكم بنفسي على ما هي سرعة القيادة المناسبة.

(كثيراً، أحياناً، نادراً، أبداً).

ولنفترض أيضًا افتراضاً غير منطقــي بأن المقياس يرتبط ارتباطاً تاماً بعدد تذاكر مخالفة السرعة التي تم الحصول عليها خلال فترة خمس ســنوات. يتم تسجيل المقياس عن طريق إعطاء كل بند القيمة (٣) عندما يختار المجيب الاستجابة «كثيراً» والقيمة (٢) لــلاستجابة «أحيانا» والقيمة (١) لــلاستجابة «نادرًا» والقيمة (صفر) لـلاستجابة «أبدًا». ثم يتم جمع درجات البنود للحصول على درجة المقياس. درجة صدق المعيار الكاملة لا تعني أن الدرجة البالغة تســع درجات تشـير إلى الحصول على تسع تذاكر مخالفة للسرعة على مدار خمس ســنوات، بل يعني هذا أن الأشــخاص الذين يســجلون أعلى الدرجات على الأداة هم أيضًا الأشــخاص الذين لديهم أكبر عدد من تذاكر مخالفة السرعة في الســنة. ومن شــأن بعض عمليات التحويل المحددة تجريبياً (مثلاً، ٠,٣٣ * الدرجة) أن تسفر عن التقدير الفعلي. من شــأن هذا التحويل المحدد أن يتنبأ بثلاث بطاقات للســائق الذي حصل على تسع درجات. وكلما زاد الصدق المرتبط بمعيار، كان التقدير أكثر دقة استناداً إلى مقياس التنبؤ. ومع ذلك، فإن التشــابه بين القيم العددية للمعيار ومقياس التنبؤ قبل القيام بعملية تحويل مناســبة لن تكون له أي علاقة بدرجة الصدق.

الصدق البنائي (صدق التكوين الفرضي):

إن الصــدق البنائي كما أشــار كرونبــاخ وميهــل (1955) Cronbach & Meehl يُعنى مباشرةً بالعلاقة النظرية لمتغير (على ســبيل المثــال، درجة على مقياس ما) بمتغيرات أخرى. وهو المدى الذي يتصرف (يســلك) فيه المقياس بالطريقة التي يفترض أن يســلكها المفهوم البنــائي الذي يســتهدف قياســه فيما يتعلــق ببناء المقاييــس للمفاهيم البنائيــة الأخرى.

لذلك على سـبيل المثـال، إذا نظرنا إلى متغير معين، اسـتنادًا إلى النظرية، على أنه مرتبط إيجابـاً بالمفهومين البنائيين A وB، ويرتبط سـلبًا بالمفهومـين البنائيين C وD، وغير مرتبط بالمفهومـين البنائيَّين X وY، فإن المقياس الذي يهـدف إلى قياس ذلك المفهوم البنائي يجب أن تكـون له علاقة مماثلة لمقاييس تلك المفاهيم البنائية الأخرى. وبعبارة أخرى، يجب أن يرتبط مقياسنا بشكل إيجابي مع مقاييس المفهومين البنائيَّين A وB، ومرتبطًا سلبًا بمقاييس المفهومـين البنائيَّين C وD، وغير مرتبط بمقاييس المفهومين البنائيَّين X وY. ويمكن التعبير عن هذه العلاقات الافتراضية كما في المخطط في الشكل رقم (٤-١).

ويوفـر مدى تطابق الارتباطات التجريبية مع النمـوذج المتوقع بعض الأدلة على مدى تصرف أو أداء المقياس كما يفعل المتغير المفترض أن يقيسه.

وبتطبيق هذا النهج العام، وصف "سـتيربا وآخـرون" Sterba et al. (2007) نمطاً من الارتباطـات التي تنبـأوا بأن تكون لأداتهم مع عدة مفاهيم بنائية ذات صلة. على سـبيل المثال، استناداً إلى تحليلهم النظري، توقعوا أن ترتبط درجات الفعالية الزوجية مع مقاييس الجـودة الزوجيـة، والتكيف النفـسي، ومعايير العمل الجماعي. وقـد حددت بعض هذه التوقعات وجود ارتباطات سـلبية، مثل العلاقة العكسـية المفترضة بـين الفعالية الزوجية والاكتئاب. وكان أحد القيود على تلك الدراسـة، التي أقرهـا المؤلفون، أن الفعالية الذاتية الفردية لم يتم تقييمها. إن إثبات أن الفعالية الزوجية تختلف عن تصورات الكفاءة الذاتية الفرديـة للشركاء داخـل الأزواج couples يُعد جانباً مهماً محتمـلاً في عملية التحقق من صدق المقياس.

	A	B	C	D	X	Y
Variable	+	+	-	-	0	0

شكل (٤-١): علاقة افتراضية بين متغيرات

التمييز بين الصدق البنائي والصدق المرتبط بمحك:

غالبًا ما يخلط الناس بين الصدق البنائي والصدق المرتبط بمحك أو معيار وذلك بسـبب أن الارتباط الدقيق نفسه يمكن أن يخدم كلا النوعين من الصدق. ويكمن الفرق في مقصد

وهــدف الباحــث أكثر مما يكمن في القيمة التي تم الحصول عليها. على ســبيل المثال، قد يحــاول أخصــائي الأوبئة تحديد أي مــن مجموعة متنوعة من المقاييــس التي تم الحصول عليها في دراسة مسحية ترتبط بالحالة أو الوضع الصحي. وقد يكون القصد أو الهدف من ذلك هو مجرد تحديد عوامل الخطر دون أي قلق (على الأقل في البداية) للآليات الســببية الأساســية التي تربط بين الدرجات على المقاييس والحالة الصحية. الصدق، في هذه الحالة، هــي الدرجــة التي يمكن أن تتنبأ بهــا المقاييس بالحالة الصحيــة. وكبديل لذلك، يمكن أن يكون الاهتمام هو النظرية والتفســير بشــكل أكبر. وقد يؤيد الباحث، مثله مثل أخصائي الأوبئــة الذي تم وصفه في الفصل الافتتاحي مــن هذا الكتاب، نموذجًا نظريًا يعتبر الإجهاد ســبباً في الحالة الصحية، وقد تكون المســألة هي مدى نجاح مقيــاس تم تطويره حديثاً في قياس الإجهاد. قد يتم تقييم ذلك من خلال تقييم ســلوك المقياس بالنســبة للكيفية التي تشــير النظريــة في الكيفية التي يجــب أن يعمل بها الإجهاد. إذا كانت النظرية تشــير إلى ضرورة الارتباط بين حالة الإجهاد والحالة الصحية، فيمكن استخدام نفس العلاقة التجريبية المستخدمة كدليل على الصدق التنبؤي في المثال السابق كدليل على الصدق البنائي.

وما يسمى بالتحقق من صدق المجموعات المعروفة known-groups validation هو مثال آخر على الإجراء الذي يمكن تصنيفه إما على أنه صدق بنائي أو الصدق المرتبط بمعيار، تبعــاً لنية وهــدف الباحث. يتضمن التحقق من صدق المجموعــات المعروفة عادةً إثبات أن مقيــاس محدد يمكن أن يميــز أعضاء مجموعة واحدة عن أعضــاء مجموعة أخرى بناءً على درجات المقياس الخاصة بهم. قد يكون الغرض إما النظرية ذات صلة (كما هو الحال عندما يتــم التحقق من قياس الاتجاهات نحو مجموعة معينة عن طريق التمييز بشــكل صحيح بين أولئك الذين ينتمون أو لا ينتمون إلى عضوية تلك المجموعة) أو التنبؤية البحتة (كما هو الحال عندما يستخدم المرء مجموعة من البنود ليس لها علاقة على ما يبدو بالتنبؤ بــترك الوظيفة). في الحالــة الأولى، ينبغي اعتبار الإجراء نوعاً من الصدق البنائي، وفي الحالة الثانية، يكون الصدق هو الصدق المرتبط بمعيار أو محك.

وبالإضافة إلى النية أو الهدف، هناك اختلاف آخر يكمن في كثير من الأحيان وراء الصدق المرتبــط بالمعيــار والصدق البنائي. غالبــاً ما يتم تقييم الصدق المرتبــط بمعيار مباشرةً عن طريق حســاب ارتباط بين المقياس الذي يتم التحقق من صدقه والمعيار (على سبيل المثال،

بعض السـلوكيات أو الحالـة أو الدرجة). وعلى النقيض من ذلـك، لا يمكن تقييم الصدق البنـائي إلا بصورة غـير مباشرة (انظر لـورد ونوفيـك Lord & Novick, 2008, p. 278)، لأن المقارنة ذات الصلة تكون بمتغير كامن وليس متغير ملاحظ. وفي هذا الصدد، فإنه يشبه الثبات. وفي الفصول السابقة، لاحظنا أن الثبات يتعلق في الحقيقة بالعلاقة بين مؤشر ودرجة حقيقية لا يمكن ملاحظتها، وأننا نسـتنتج تلك العلاقة على أساس الارتباطات بين المؤشرات التي يمكن ملاحظتها. وينطبق نفس الشيء على الصدق البنائي. حيث لا يمكننا أن نحسب مباشرة الارتباط بين أداة القياس الذي يتم التحقق من صدقها والمتغير الكامن، ولكن يجب علينـا أن نفعـل ذلك بصورة غير مباشرة مـن خلال ملاحظة الارتباطات بـين أداة القياس الجديـدة والمؤشرات الأخرى الموثوق بها للمتغير الكامن. وفي الواقع، قد يجد الباحثون أنه من المفيد اسـتخدام هذا الاختلاف أو الفرق كقاعدة حسـابية تقريبية عندما يكونون غير متأكدين مما إذا كان الصدق المرتبط بمعيار أو الصدق البنائي هو موضع الاهتمام الأساسي. إذا كان الهدف هو التنبؤ بمخرجات أو نتائج يمكن ملاحظتها (على سـبيل المثال، السـلوك، أو الحالـة، أو الدرجة الملاحظة)، فإن الصدق المرتبط بمعيار أو محك قد يكون هو الهدف. ومن ناحية أخرى، إذا كان الهدف هو التنبؤ بمستوى مفهوم بنائي افتراضي، لا يمكن ملاحظته عن طريق مؤشر يمكن ملاحظته، فمن المرجح أن يكون الهدف هو الصدق البنائي.

ومع ذلك، في بعض الأحيان، سـيتضمن التحقق من الصدق المرتبط بمعيار مؤشراً يُقصد به أن يعكس أو يظهر بعض الدرجات الحقيقية غير الملاحظة. وهذا هو الحال عندما يكون المعيار (على سـبيل المثال، خريج كلية الحقوق المؤهل للقبول في نقابة المحامين) قائماً على شـكل ما من أشكال التقييم النفسي (على سـبيل المثال، اجتياز امتحان المحاماة). إذا ظل هـدف الباحـث هو مجرد تنبؤ بدون تفصيل مفاهيمي، فإن مثـل هذه المقارنة هي مثال على الصدق المرتبط بمعيار على الرغم من عدم المقارنة بين المتنبئ والنتائج موضع الاهتمام.

التخفيف (التضعيف) Attenuation:

إلى الحد أو القدر الذي يكون فيه مؤشران اثنان لا يتمتعان بالثبات التام، فإن أي ارتباط بينهما سـوف يقلل من تقدير الارتبـاط بين الدرجات الحقيقيـة المقابلة لكل منهما (على سـبيل المثال، لورد ونوفيك Lord & Novick, 2008). وعندما نفترض أن الخطأ عشـوائي،

فإنــه لا يمكن الربط إلا بين الأجزاء التي تتمتــع بالثبات من كل من المؤشرين فقط. وبناءً على ذلك، يضعف (ينخفض) معامل الصدق الملاحظ المحســوب بالربط بين متغيرين يمكن ملاحظتهــما كنتيجة لأي عــدم ثبات متأصــل في تلك المؤشرات. ويمكن للشــخص تطبيق تصحيح الضعف من علاقة ارتباط ملحوظ تأخذ في الاعتبار عدم ثبات المتغيرات. وينطوي هذا التصحيح على تقسيم الارتباط الملاحظ بين المؤشرين على الجذر التربيعي لحاصل ضرب ثبات كل منهم. وهكذا،

$$r(T)_{xy} = \frac{r_{xy}}{\sqrt{r_{xx} r_{yy}}}$$

حيــث $r(T)_{xy}$ هــي الارتباط بين الدرجــات الحقيقية للمتغــيرات X وY، و(r_{xy}) هي العلاقة بين الدرجات الملاحظة، و(r_{xx}) و(r_{yy}) هي ثبات المتغيرات X وY، على التوالي.

وعلى الرغم من توافر مثل هذه التصحيحات، فإن بعض المؤلفين أو الكتاب يعارضون هذه التصحيحات. فعلى سبيل المثال، يشير "نونالي وبيرنشتاين" Nunnally and Bernstein (1994) إلى أنها قد تضلل الباحثين إلى الاعتقاد بأن الارتباط أقوى مما هو عليه في الواقع. كما يلاحظون أن المعامــلات التي تم تصحيحها يمكــن أن تتجاوز أحيانًا قيمة الواحــد الصحيح (١,٠) (ص ٢٥٧)، وهذا يمثل مشــكلة. ويشــير لورد ونوفيك Lord and Novick (2008) إلى أنه عندما يتم التقليل من تقدير ثبات كل من المقياسين (والتي تحدث في مقام معادلة التصحيح)، فإن الارتبــاط بين الدرجات الحقيقية ســيكون مبالغاً في تقديره (ص ١٣٨). وبالإشــارة إلى ما تم تناوله في الفصول الســابقة من أنه وفقًا لنــماذج أخرى غير الاختبارات المتكافئة، فإن معامل «ألفا» يكون تقديراً للحد الأدنى للثبات الحقيقي للمقياس، يصبح من الواضح أن ينشأ بسهولة «تصحيــح» غير دقيق. وبالتالي، هناك حجج عملية قوية ضد اســتخدام التصحيح للتضعيف عند فحص الارتباطات، سواء من أجل تقييم الصدق أو لأغراض أخرى. وبسبب الطبيعة غير المباشرة لمعاملات الارتباط المســتخدمة في تقييم الصدق البنائي، كما تمت مناقشــته في وقت سابق، فقد يكون من المغري بشكل خاص تصحيح معاملات الصدق من أجل التضعيف. ومع ذلك، لا تزال الحجج ضد تعديل التضعيف منطبقة، ومن الممارســات الشــائعة عدم تصحيح معامل الارتباط في مثل هذه الحالات.

إلى أي مدى يجب أن تكون الارتباطات قوية لإظهار الصدق البنائي؟

لا توجد نقطة قطع تحدد الصدق البنائي. ومن المهم أن ندرك أن هناك مقياسين قد يتشاركان في أكثر من مفهوم بنائي متشابه. على وجه التحديد، فإن أوجه التشابه في الطريقة التي يتم بها قياس المفاهيم البنائية قد تفسر بعض التغاير في الدرجات المستقلة عن التشابه في المفهوم البنائي. على سبيل المثال، فإن متغيرين اثنين تم تسجيلهما على نظام تسجيل متعدد النقاط (الدرجات من ١ إلى ١٠٠) سيكون لهما ارتباط أعلى مع بعضهما البعض مقارنة بمتغير ثنائي، مع تساوي كل شيء آخر. وهذا يكون خادعاً بسبب هيكل أو بناء طرق القياس. وبالمثل، نظرًا للتشابه الإجرائي، قد ترتبط البيانات من نوع واحد تم جمعها عن طريق المقابلات بدرجة ما مع البيانات الأخرى التي تم جمعها بنفس الطريقة - أي إن بعض التغاير بين متغيرين قد يكون بسبب تشابه القياس بدلاً من تشابه المفهوم البنائي. وتوفر هذه الحقيقة الأساس للإجابة على السؤال المتعلق بحجم الارتباطات اللازمة لاستنتاج الصدق البنائي. وينبغي أن تظهر المتغيرات، كحد أدنى، التغاير أعلى وأكثر مما يمكن أن يعزى إلى تباين الطريقة المشتركة.

مصفوفة الصفات المتعددة - الأساليب المتعددة Multitrait-Multimethod matrix:

ابتكر «كامبل وفيسكي» Campbell and Fiske (1959) إجراءً يسمى مصفوفة الصفات المتعددة - الأساليب المتعددة multitrait-multimethod matrix التي تُعد مفيدة للغاية لفحص ودراسة الصدق البنائي. ويتفق هذا النهج أو الطريقة تماماً مع الفكرة، التي تم تقديمها مسبقاً، التي ترى أن الصدق البنائي يتم تقييمه بصورة غير مباشرة ويجب الاستدلال عليه من المؤشرات المتاحة للمتغير الكامن محل الاهتمام. وينطوي الإجراء على قياس أكثر من مفهوم بنائي واحد بأكثر من طريقة بحيث يتم الحصول على مصفوفة «متقاطعة تماماً» حسب مصفوفة القياس. على سبيل المثال، لنفترض أنه قد تم تصميم دراسة ما يتم فيها قياس كل من القلق والاكتئاب ومقاس أو حجم الحذاء في مرتين منفصلتين باستخدام إجراءين قياسيَّين مختلفين في كل مرة. (لاحظ أنه كان من الممكن قياس عينتين مختلفتين من الأفراد في نفس الوقت. ما هو تأثير ذلك على منطق النهج أو الطريقة؟) ويمكن تقييم

كل مفهوم بنائي بطريقتين: مقياس تناظري بصري (خط أو سطر يضع عليه المجيبون علامة تشير إلى مقدار السمة التي يمتلكونها، سواء كانت القلق أو الاكتئاب أو كبر حجم القدم) وتصنيف معين من قبل الشخص الذي يجري المقابلة بعد ١٥ دقيقة من التفاعل مع كل موضوع. يمكن للمرء بعد ذلك بناء مصفوفة من الارتباطات التي تم الحصول عليها بين القياسات كما هو مبين في الجدول رقم (٤-١).

هناك تمييز آخر محتمل، غير موجود في الجدول، وهو بين السمات ذات الصلة مقابل السمات غير ذات الصلة. لأن المدخلات «البنود» التي تعكس نفس السمة (المفهوم البنائي) ونفس الطريقة ينبغي أن تشترك في كل من تباين الأسلوب «الطريقة» وتباين المفهوم البنائي، فمن المتوقع أن تكون هذه الارتباطات أعلى. ومن المؤمل أن تكون الارتباطات المقابلة لنفس السمة ولكن الأساليب مختلفة هي التالية في العلو. إذا كان الأمر كذلك، فإن هذا يشير إلى أن التغاير للمفهوم البنائي أعلى من التغاير للأسلوب أو الطريقة. وبعبارة أخرى، إن مقاييسنا كانت أكثر تأثراً بما تم قياسه من كيفية أو طريقة قياسه. وعلى النقيض من ذلك، لا يوجد سبب يدعو إلى وجود أي تغاير بين حجم الحذاء وأي من المفهومين البنائيَّين الآخرين عندما يتم قياسهما بإجراءات مختلفة. ومن ثم، يجب ألا تختلف هذه الارتباطات اختلافًا كبيرًا عن الصفر. بالنسبة للمفاهيم البنائية غير المتطابقة ولكنها مرتبطة من الناحية النظرية، مثل الاكتئاب والقلق، قد يتوقع الشخص وجود بعض التغاير للمفهوم البنائي.

ومن المحتمل أن تكون هذه مجموعة معلومات مفيدة للغاية من الارتباطات المتبادلة لإيجاد وإثبات الصدق البنائي. على سبيل المثال، إذا كانت مقاييس الاكتئاب لدينا راسخة ومستقرة جيداً على حد سواء ولكن يجري حالياً وضع وتطوير مقاييس القلق لدينا، يمكننا تقييم مقدار التغاير الذي يعزى إلى تشابه المفاهيم في ظل إجراءات قياس مماثلة ومختلفة. تؤكد النظرية على أن القلق والاكتئاب ينبغي أن يكونا مرتبطين إلى حد كبير حتى لو تم قياسهما بطرق مختلفة. وإذا ثبت أن هذا هو الحال، فإنه سيكون بمثابة دليل على الصدق البنائي لمقياس القلق الجديد لدينا. وبشكل أكثر تحديداً، فإن هذه الارتباطات تشير إلى الصدق التقاربي convergent validity، أي دليلاً على التشابه بين مقاييس المفاهيم البنائية المرتبطة نظرياً. ومن الناحية المثالية، فإن الارتباطات بين القلق والاكتئاب ستكون أقل من تلك التي توجد بين اثنين من مقاييس الاكتئاب أو اثنين من مقاييس القلق ولكن

أكـبر بكثير من بـين أي من درجات الاكتئاب وحجم الأحذيـة. وبنفس القدر من الأهمية دليـل على أن مقاييس القلق لم ترتبـط ارتباطاً ذا دلالة إحصائيـة بمقاييس حجم الحذاء، بغـض النظر عن التشـابه أو الاختلاف في تقنية القياس. هـذا دليل على الصدق التمايزي discriminant validity (ويسـمى أحياناً الصدق المتبايـن divergent validity)، وعدم وجـود ارتباط بين مقاييس المفاهيـم البنائية غير ذات الصلة. يشـير الارتباط ذي الدلالة الإحصائية بين حجم الحذاء والقلق عند قياسها بنفس الطريقة إلى أن الطريقة في حد ذاتها تمثل مقداراً كبيراً من التباين (والتغاير) المرتبط بمقاييس مماثلة للمفاهيم البنائية المختلفة.

جدول (٤-١) تفسيرات الارتباطات في مصفوفة الصفات المتعددة - الأساليب المتعددة

		Time 1					
Time 2		A_v	A_i	D_v	D_i	S_v	S_i
	A_v	TM	T	M		M	
	A_i	T	TM		M		M
	D_v	M		TM	T	M	
	D_i		M	T	TM		M
	S_v	M		M		TM	T
	S_i		M		M	T	TM

ملاحظات: (TM) تساوي نفس السمة ونفس الطريقة أو الأسلوب (الثبات)؛ و(T) نفس السمة، وطريقة مختلفة؛ و(A) و(D) و(S) تشـير إلى التكوينات الفرضية القلق، والاكتئاب، وحجم الأحذية على الترتيب. الرموز السـفلية (v) و(i) تشير إلى أساليب تناظرية بصرية وطرق المقابلة، على التوالي.

ولاحـظ ميتشـل Mitchell (1979) أن الأسـاليب والطرق التي تنطـوي عليها عملية جمـع البيانات لمصفوفة الصفات المتعددة - الأسـاليب المتعددة تشـكل دراسـة إمكانية التعميـم ذات وجهـين أو جانبـين two-facet G-study (أو دراسـة قابليـة التعميـم؛ انظـر الفصـل الثالث)، مـع اعتبار أن السـمات والأسـاليب هـي الأوجـه أو الجوانب.

وتتيــح لنــا مصفوفة الصفات المتعددة - الأســاليب المتعددة تقســيم التغاير إلى مصادر «الأسلوب» ومصادر «السمة» (أو المفهوم البنائي). ويمكننا بعد ذلك تقديم بيانات أكثر دقة حول الصدق البنائي، لأنها تســمح لنا بالتمييز بين التغاير الذي يعكس بالفعل التشــابه بين المفهــوم البنائي (وبالتالي يرتبط بالصدق البنائي) عن التغاير الذي هو نتاج تطبيق إجراءات قياس متشــابهة (وبالتالي لا يرتبط بالصدق البنائي). ولا يمكــن تحقيق هذا التمييز عندما ينظر الشخص ببساطة في ارتباط واحد بين اثنين من المقاييس.

ماذا عن الصدق الظاهري؟

يستخدم العديد من الأشخاص مصطلح «الصدق الظاهري» لوصف مجموعة من البنود التي تقيم ما يبدو أنها تقيسه. في رأيي، هذا الاستخدام سيئ لعدة أسباب:

أولاً، الافتراض بأن المقياس يقوم بتقييم ما يبدو أنه يقيسه قد يكون خاطئًا. فعلى سبيل المثال، قام كل من «إدلر وبنياميني» Idler and Benyamini (1997) بفحص ٢٧ دراســة وبائية كبيرة وأجريت بطريقة جيدة لتحديد ما هو البند المشترك الذي تمت الاستفادة منه على وجه الدقة. يطلب هذا البند من الناس تقييم صحتهم بشــكل عام على أنها إما سيئة أو متوسطة أو جيدة أو ممتازة. معظم الناس سيحكمون على هذا المقياس المكون من بند واحد لتقييم ما يقولونه بالضبط: صحة المستجيبين. وقد لاحظ "إدلر وبنياميني" أن البند كان مؤشراً ممتازاً لمجموعة متنوعة من النتائج الصحية. وقد تفوق أداء هذا البند باستمرار على المتغيرات الأخرى في حســاب التباين عبر الدراسات المختلفة. والأكثر ارتباطاً بمناقشتنا، يبدو أنه لا يتعلق بشــكل أســاسي بالحالة الصحية. وكثيراً ما تتضمــن النماذج بنداً واحداً وتضــع أيضاً مقاييــس للحالة الصحية. وعادةً ما تكون مقاييــس التقييم الذاتي ذات البند الواحــد للحالة الصحية والمقاييس الأخــرى للحالة الصحية منبِّئَين هامين في نفس النموذج. أي إنها لم تشترك في التباين الكافي من أجل المساهمة التنبؤية من جانب واحد لمنع مساهمة تنبؤية مستقلة من الآخر. بل يبدو أن التقييم الذاتي ذا البند الواحد للحالة الصحية يشترك في التباين بدرجة أكبر مع المتغيرات النفســية. وتشير هذه النتائج إلى أن هذا البند الواحد المســتخدم على نطاق واســع ليس مؤشرًا صادقاً للحالة الصحية، كما يظهر. بالنســبة لهذا البند، فإن النظر إلى ما إذا كان يقيس ما نريد قياسه ليس كافياً لدعم ادعاءات الصدق.

والمشــكلة الثانية في تقييم المقياس اســتناداً إلى الصدق الظاهري هــي أن هناك أوقاتاً يكون فيها من المهم ألا يكون المتغير الذي يجري قياســه واضحاً. فعلى ســبيل المثال، فإن أداة ما تهدف إلى تقييم الدرجة التي يجيب بها الأشخاص على نحو غير صحيح (كذباً) (على ســبيل المثال، لجعل أنفسهم "يبدون جيدين") لن يكون من المفيد أن يكون غرضه واضحاً للمجيبين. هل نســتنتج أنه غير صحيح لأنه لا يبدو كما لو كان يقيس عدم الصدق؟ أتمنى ألا يكــون كذلك. وهكــذا، لدينا هنا حالة لا يمكن فيها أن يؤيد عدم النظر إلى ما هو عليه في الواقع استنتاج البطلان.

ومن الشواغل والاعتبارات الأخيرة فيما يتعلق بالصدق الظاهري أنه ليس من الواضح لمن ينبغي أن يكون الغرض من الأداة واضحاً، في ظاهرها. هل هو الـمستجيب؟ إذا سأل الطبيب المريض عما إذا كان هو أو هي أكثر عطشاً من المعتاد، فهل يتوقف صدق هذا السؤال على معرفة المريض بما طُلب منه أو سُــئل عنه؟ من الواضح أنه ليس كذلك. هل الشــخص الذي يصمــم ويبني الأداة هو الذي ينبغي أن يدرك ويعــرف الغرض من الأداة؟ ومن الصعب أن نتصور أن الصلة بين محتوى الأداة والمتغير موضع الاهتمام ليســت واضحة بالنســبة لمنشئ أداة مــا (ربما باســتثناء الحالات التي تتعلق بالصدق التجريبــي البحت، والصدق اللانظري atheoretical، والصــدق المرتبط بمحــك). وإذا تم اعتماد هذا المعنــى للصدق الظاهري، فســيتم اعتبار جميع المقاييس صادقة أساســاً. وأخيراً، هل هو مجتمع علمي أكبر أو أوسع ينبغي أن يدرك ويعرف غرض أداة ما على أساس مظهرها (ظاهرها)؟ ومن المرجح أن يسفر هذا التفســير عن أدلة متضاربة. قد يبدو بند ما إلى بعض الخبراء كما لو كان يقيس متغيراً واحــداً وكأنه يقيس متغيراً آخر من وجهة نظر مجموعة خبراء أخرى مؤهلة بنفس الدرجة. وغالباً ما يبدو أن الأشخاص الذين يدّعون أن مقياس ما يتمتع بالصدق أو لا يتمتع بالصدق لأنــه يبــدو أنه يتمتع بالصدق الظاهــري أو لا يتمتع بالصدق الظاهري تســتند ادعاءاتهم إلى تصوراتهم الشــخصية. أي إنه إذا كان القصد من أداة ما ومظهرها متشــابهان لبعضهما البعض، فإنهم يميلون إلى اعتبار أن الأداة تتمتع بالصدق الظاهري؛ خلاف ذلك، فهي ليست كذلك. ويبدو أن ذلك يشكل أساساً ضعيفاً لأي ادعاء بالصدق.

يتم الخلط أحيانًا بين الصدق الظاهري وصدق المحتوى نظرًا لأن كليهما قد يتعلق بمدى مـا يبدو عليــه محتوى البند ذي الصلــة بالمفهوم البنائي محل الاهتــمام. ومع ذلك، هناك اختــلاف مهم بينهما يتمثل في أن صدق المحتــوى يتم تحديده من ناحية إجراءات محددة، وهذه الإجراءات بشكل عام تكون أكثر تنظيماً ودقة وصرامة مقارنة بالتقييمات غير الرسمية للصــدق الظاهــري. وكما لوحظ، فقد تتضمــن هذه الخطوات جمع الأفــكار من المجيبين المحتملين، كما تشــمل عادةً الحصول على معلومات من الأدبيات النظرية ذات الصلة ومن الخبراء في المجال قيد البحث. على سبيل المثال، قد يُطلب من خبراء المحتوى تقييم محتوى البنــد في علاقته بتعريــف مفهوم بنائي صريح وواضح. ومن المفــترض أن يكون لدى هؤلاء الخــبراء إطار مرجعي نظري يدعم أحكامهم. فمجرد ظهــور الصلة أو العلاقة لا يكون هو المعيار الوحيد. كما أن الأفراد الذين تُســتخدم أحكامهــم لتقييم الصدق يكونون واضحين. عــلاوة على ذلك، فإن الاســتنتاج بأن الأداة تتمتــع بصدق المحتــوى لا يقتصر على تحديد الباحثين أنفسهم. وأخيرًا، فإن شفافية المجيب ليست أساسًا لتقييم صدق المحتوى لأنه يتم تقييمه عادةً (على ســبيل المثال، من خلال أحكام الخبراء). وبالتالي، فإن الطريقة الرســمية للتحقق من المحتوى يمكن أن تتغلب على قيود ومحددات الاعتماد على الصدق الظاهري.

وتبعــاً للظروف والأحوال، قد تكون هناك مزايا أو عيوب في أن يكون الهدف من الأداة واضحاً من مظهره. وكما ســنرى في الفصل التالي، غالبًا ما تنتج عملية تكوين وتوليد البنود عبارات تشــير صراحة إلى المتغير محل الاهتمام، وهذا عادةً ليس أمراً سيئًا. أنا لا أقترح أنه يجب بناء الأدوات بشــكل عام بحيث لا يكون القصد أو الهدف منها واضحًا من مظهرها؛ بدلاً من ذلك، فإنني أقترح أنه سواء كان الأمر كذلك أم لا، فليس له علاقة بموضوع الصدق.

تمارين:

١- اذكر مثالاً على كيف أن نفس الارتباط بين مقياس ما وسلوك ما قد يكون مؤشراً أو دلالة عــلى كل من الصــدق البنائي أو الصــدق المرتبط بمعيار «محــك»، واشرح كيف أن (أ) الدوافع وراء حســاب الارتباط، و(ب) تفســير هذا الارتباط يختلف تبعاً لنوع الصدق الذي يحاول الباحث تقييمه.

٢- افـترض أن لدى الباحث مقاييس ورقية وقلـم رصاص لقياس اثنين من المفاهيم البنائية هما: احترام أو تقدير الذات والتوافق الاجتماعي. الباحث لديه أيضاً درجات اسـتناداً على مقابلات شخصية على نفس المفهومين البنائيَّين. كيف يمكن استخدام هذه البيانات في مصفوفة الصفات المتعددة - الأساليب المتعددة لإظهار أن طريقة جمع البيانات كان لها تأثير قوي غير مرغوب فيه على النتائج التي تم الحصول عليها؟

٣- كيف يمكن أن يؤثر تعريف نطاق أو مجال المحتوى بشـكل واسـع للغاية لمقياس ما تأثيراً سـلبياً على صدق المحتوى الخاص بهذا المقياس عند اسـتخدامه مع مجتمعات إحصائية معينة؟

الفصل الخامس
قواعد إرشادية في تطوير المقياس

حتى الآن، كانت مناقشـتنا مجردة إلى حدٍ ما. ونحن الآن ننظر في كيفية تطبيق هذه المعرفـة. ويقدم هذا الفصل مجموعة من المبادئ التوجيهية أو الإرشـادات المحددة التي يمكن للباحثين استخدامها في وضع وتطوير القياس والمقياس.

الخطوة الأولى: حدد بوضوح ما تريد قياسه:

يعتقد العديد من الباحثين أن لديهم فكرة واضحة عما يريدون قياسـه فقط ليجدوا أن أفكارهـم أكثر غموضاً مما كانوا يعتقدون، وهذا واضح بشـكل خادع أو مضلل. وكثيراً ما يحدث هذا الإدراك بعد أن تم بذل جهد كبير في إنشـاء وتوليد البنود وجمع البيانات - في الوقـت الـذي تكون فيه التغييرات أكثر تكلفة بكثير مما لو تم اكتشـافه في بداية العملية. هل يجب أن يكون المقياس قائماً على نظرية، أم يجب أن تبدأ في اتجاهات فكرية جديدة؟ لأي درجـة من الدقة ينبغي أن يكون عليها المقياس؟ هل ينبغي التأكيد على بعض جوانب الظاهرة أكثر من جوانبها الأخرى؟

النظرية كوسيلة مساعدة على تحقيق الوضوح:

كـما هو موضح في الفصل الأول، يتطلب التفكـير بوضوح في محتوى مقياس ما التفكير جليـاً في المفهوم البنائي الذي يتم قياسـه. وعلى الرغم مـن أن هناك العديد من الجوانب التقنيـة التي ينطوي عليها وضع وتطوير مقياس والتحقق من صدقه، فإنه ينبغي ألا نغفل أهمية أن يكون له أسـاس سليم في النظريات الأساسـية والموضوعية ذات الصلة بالظاهرة المراد قياسها. والغرض من أنواع المقاييس التي هي محور التركيز الرئيسي لهذا الكتاب هو قيـاس الظواهر المراوغة التي لا يمكن ملاحظتها مباشرة. ونظرًا لعدم وجود معيار ملموس يمكـن على أساسـه مقارنة أداء هذا النوع من المقاييس، فمـن المهم أن يكون لديك بعض الأفكار الواضحة لتكون بمثابة دليل. ويجب إدراك وتحديد حدود الظاهرة حتى لا ينجرف محتوى المقياس دون قصد إلى مجالات غير مقصودة.

تُعد النظرية مساعداً كبيراً على تحقيق الوضوح. وينبغي دائماً الأخذ في الاعتبار نظريات العلــوم الاجتماعيــة ذات الصلة قبل وضع وتطوير مقياس من النوع الذي تمت مناقشــته في هــذا الكتــاب. وإذا تبــين أن النظرية الموجــودة لا تقدم أي إرشــاد أو توجيه لواضعي أو مطــوري المقاييس، فقد يقررون حينئذ أن الاتجــاه الفكري الجديد ضرورياً. ومع ذلك، ينبغي أن يكون هذا القرار مستنيرًا ومدروساً، ولا يتم التوصل إليه إلا بعد مراجعة النظرية المناسبة المتعلقة بمشكلة القياس قيد النظر. وحتى لو لم تكن هناك نظرية متوفرة لتوجيه الباحثين، فيجب عليهم أن يضعوا صيغتهم المفاهيمية الخاصة قبل محاولة قياسها إجرائياً. وينبغي على الباحثين، في الأساس، أن يحددوا على الأقل نموذجًا نظريًا أولياً أو مبدئيًا يكون بمثابة مرشــد أو دليل لوضع وتطوير المقياس. وقد يكون هذا النموذج الأولي بســيطًا مثل مجرد تحديد الصياغة الجيدة والموضوعية للظاهرة التي يســعون إلى قياسها. والأفضل من ذلك إدراج وصف لكيفية ارتباط البناء المفاهيمي الجديد بالظواهر القائمة وقياسها إجرائياً.

الخصوصية كوسيلة مساعدة على الوضوح:

وقد يكون مســتوى التحديد «الخصوصيــة» أو العمومية التي يتــم بها قياس المفهوم البنــائي أمراً مهماً أيضــاً. وهناك اتفاق عام في العلوم الاجتماعية على أن المتغيرات ســوف ترتبط بقوة مع بعضها البعض عندما تتطابق في مستوى التحديد أو الخصوصية (انظر أجزن وفيشــبن Ajzen& Fishbein,1980، للاطلاع على مناقشة ذلك). وفي بعض الأحيان، يكون الهدف من المقياس أن يرتبط أو يتعلق بســلوكيات أو مفاهيم بنائية محددة للغاية، بينما في أحيان أخرى، يكون المطلوب مقياس أكثر عمومية وشمولية.

كمثــال توضيحي للمقاييس التي تختلف في درجة التحديد، نأخذ بعين الاعتبار المفهوم البنــائي لموضــع التحكم أو الضبط. موقع التحكم هو مفهوم يســتخدم على نطاق واســع ويتعلق بتصورات الأفراد حول من أو ما يؤثر على النتائج المهمة في حياتهم. ويمكن تطبيق هذا المفهوم البنائي على نطاق واســع، كوســيلة لشرح أنماط الســلوك الشــامل أو العالمي التــي تغطي العديد من الحــالات والمواقف، أو على نطاق ضيق، للتنبؤ بكيفية اســتجابة الفرد في ســياق محدد للغايــة. ويمكن أيضاً وصف مصادر التأثير إما على نطاق واســع أو عــلى نطاق ضيق ومحدد للغاية. وعلى ســبيل المثال، فإن مقياس "روتر" Rotter (1966)

لموضــع التحكــم أو الضبط الداخــلي - الخارجي، يهتم بالقياس على مســتوى عام إلى حد مــا لهــذه التصورات. وثمة بُعد واحــد يتراوح بين التحكم الشــخصي إلى التحكم من خلال عوامل خارجية يُعد هو الأســاس الذي يقوم عليه المقياس، والنتائج التي تركز عليها البنود تكون عامة، مثل النجاح الشــخصي. كما أن المصادر الخارجية للتحكم أو السيطرة وُصفت بعبــارات عامة أيضاً. العبارات التالية خاصــة بموضع التحكم الخارجي من مقياس "روتر" لموضع التحكم الداخلي - الخارجي: "إن العالم يديره عدد قليل من الأشــخاص الموجودين في السلطة، وليس هناك الكثير من الشباب الصغير يمكنهم القيام بذلك".

وقــد طــورت ليفنســون Levenson (1973) مقيــاس متعدد الأبعــاد لموضع التحكم أو الضبط يســمح بوجود ثلاثة مواضع للتحكم هي: الذات، والأشــخاص الأقوياء الآخرون، والفرصة أو المصير. وهذا يتيح للباحث أن ينظر إلى مصادر التحكم الخارجية بشــكل أكثر تحديداً من خلال وصفها بأنها إما الأشــخاص الآخرون الأكثر قوة أو المصير. غير أن النتائج التــي ركزت عليهــا ظلت عامة. مثال على بند من مقياس «ليفنســون» ‹للمقياس الفرعي الآخرون الأقوياء هو «أشعر أن ما يحدث في حياتي يتم تحديده من قبل الآخرين الأقوياء».

وقــد طــور والســتون، ووالســتون، ودي فيلبيــس Wallston, Wallston, and DeVellis (1978) مقاييس موضع التحكم في الصحة متعددة الأبعاد (MHLC) باستخدام المواضــع الثلاثة للتحكم التي اقترحتها "ليفنســون"، مع نتائج خاصــة بالصحة، مثل تجنب المــرض أو الإصابــة بالمرض. وهذا بند كعينة من بنود مقيــاس الآخرين الأقوياء من مقياس موضــع التحكم في الصحــة متعددة الأبعاد MFLC هو "أن يكون لــدي اتصال منتظم مع طبيبــي هــو أفضل طريقة لتفادي المرض". وبعد ذلك طور والســتون، وســتين، وســميث Wallston, Stein, and Smith (1994) مقياساً أكثر تحديداً لنتائج قياس موضع التحكم في الصحة سُمي ب (MHLC Form C) والذي يتكون من سلسلة من البنود " نموذج". ويتيح هذا الإجراء للباحث تحديد أي مشــكلة صحية ذات أهمية عن طريق اســتبدال اسم المرض أو الاضطــراب بعبارة "حالتي" في كل بند من بنــود النموذج. إن أحد البنود كعينة لمقياس الآخرين الأقوياء من مقياس (MHLC Form C)، حيث يمكن اســتخدامه في دراســة مرض السكري، هو "إذا رأيت (زرت) طبيبي بانتظام، فأنا أقل عرضة للإصابة بمرض السكري".

ومن المحتمل أن يكون كل مقياس من مقاييس موضع التحكم التي تكون أكثر تحديداً بصورة تدريجية مفيدة. والأمر الأكثر فائدة يعتمد إلى حد كبير على مستوى النتيجة أو عمومية الموضع المتعلق بالسؤال العلمي الذي يتم طرحه. على سبيل المثال، إذا كان المقصود أو الهدف من مقياس موضع التحكم هو التنبؤ بفئة عامة من السلوك أو ستتم مقارنتها مع متغيرات أخرى لتقييم مفاهيم بنائية على المستوى العام، فقد يكون مقياس "روتر" Rotter هو الخيار الأفضل لأنه هو مقياس عام أيضاً. من ناحية أخرى، إذا كان الباحث مهتماً بالتنبؤ على وجه التحديد بكيفية تأثير المعتقدات حول تأثير أشخاص آخرين على بعض السلوكيات الصحية، فإن مقياس (والستون وآخرون، ١٩٩٤) قد يكون أكثر ملاءمة لأن مستوى التحديد يتطابق مع سؤال البحث. وخلال عملية وضع وتطوير كل من هذه المقاييس، كان لكل من هذه المقاييس إطار مرجعي واضح يحدد مستوى التحديد المناسب، بالنظر إلى الوظيفة المقصودة للمقياس. والنقطة المهمة هي أن مطوري المقياس يجب أن يتخذوا قرار التحديد كقرار نشط وليس مجرد توليد مجموعة من البنود، ومن ثم رؤية الشكل الذي تبدو عليه البنود بعد الحقيقة.

يوضح المثال الخاص بمقياس موضع التحكم الخصوصية أو التحديد فيما يتعلق بالنتائج (على سبيل المثال، كيفية إدارة العالم مقابل مشاكل مرض السكري) وموضع التحكم (أي خارجي بشكل عام مقابل المصير والآخرين الأقوياء بشكل منفصل). ومع ذلك، يمكن أن تختلف خصوصية وتحديد المقياس على طول عدد من الأبعاد، بما في ذلك مجالات أو نطاقات المحتوى (على سبيل المثال، القلق مقابل التكيف النفسي على نطاق أوسع)، أو الإعداد (على سبيل المثال، الاستبيانات المصممة خصيصاً من أجل ملاءمة بيئات عمل معينة)، أو المجتمع الإحصائي (على سبيل المثال، الأطفال مقابل البالغين أو الأفراد العسكريين مقابل طلاب الجامعات).

توضيح ما يجب تضمينه في القياس:

يجب على واضعي ومطوري المقياس أن يسألوا أنفسهم عما إذا كان المفهوم البنائي الذي يرغبون في قياسه مختلفاً ومتميزاً عن غيره من المفاهيم البنائية الأخرى. وكما ذُكر آنفاً، يمكن وضع مقاييس بحيث تكون واسعة نسبياً أو ضيقة نسبياً فيما يتعلق بالحالات الذي تنطبق

عليهــا. وهذا هو الحــال أيضاً فيما يتعلق بالمفاهيم البنائية التي تغطيها. يُعد قياس القلق العــام أمــراً مشروعاً تماماً. وقد يقيس مثل هذا المقيــاس كلاً من القلق من الاختبار والقلق الاجتماعــي على حد ســواء. وهذا يُعد أمــراً جيداً إذا كان مطابقاً لأهــداف واضع (مطور) المقياس أو مستخدم المقياس. ومع ذلك، إذا كان الشخص مهتماً بنوع واحد فقط محدد من القلق، فيجب في هذه الحالة أن يستبعد المقياس جميع الأنواع الأخرى للقلق. يمكن للبنود التي قد «تتقاطع أو تشترك» في مفهوم بنائي ذي صلة (على سبيل المثال، الاستفادة من القلق الاجتماعي عندما يكون موضوع الاهتمام هو قلق الاختبار) أن تكون إشكالية.

في بعــض الأحيان، قد تتم الاســتفادة من (تأثير) البنود المتشــابهة على ما يبدو في قياس مفاهيــم بنائية مختلفة تماماً. في مثل هذه الحــالات، على الرغم من أن الغرض من المقياس قــد يكون قياس ظاهرة واحدة، فإنه قد يكون أيضاً حساســاً (مناســباً) لظواهر أخرى. على ســبيل المثال، بعض مقاييس الاكتئاب، مثل مقياس الاكتئاب الخاص بمركز الدراسات الوبائية «رادلــوف» (Radloff, 1977)، تحتوي على بعض البنود التي تســتخدم الجوانب الجســدية للاكتئاب (على ســبيل المثــال، فيما يتعلق بقــدرة المجيب على "الانطــلاق والتحرك"). وفي ســياق بعــض الظــروف الصحية، مثل التهــاب المفاصل، قد تــؤدي هذه البنــود إلى أخطاء في جوانــب المرض بســبب أعــراض الاكتئــاب (انظر بلالــوك، ديفيليس، براون، ووالســتون Blalock, DeVellis, Brown, &Wallston, 1989، لمناقشــة هــذه النقطــة تحديداً). قد يختار الباحث الذي يطور مقياسًا جديدًا للاكتئاب تجنب الجوانب الجسدية إذا كان من المقرر اســتخدام المقياس مع مجتمعات إحصائية معينة (مثل المرضى بمرض مزمن) أو مع مقاييس أخرى للمفاهيم البنائية الجســدية (مثل قصور الغضروف الدرقي "الوسواس القهري"). وقد يكون من المهم جداً، بطبيعة الحال، أن يُستخدم لأغراض أخرى إدراج وتضمين بنود جسدية، كما هو الحال عندما يتعلق البحث على وجه التحديد بالجوانب الجسدية للتأثير السلبي.

الخطوة الثانية: توليد مجموعة من البنود:

بمجرد أن يتم توضيح الغرض من المقياس بصورة محددة، يكون واضع أو مطور المقياس جاهزًا للبدء في بناء وإنشــاء أداة القياس بشــكل جدي. وتتمثل الخطوة الأولى في إنشــاء مجموعة كبيرة من البنود المرشحة للإدراج والتضمين بشكل نهائي في المقياس.

اختر البنود التي تعكس الغرض من المقياس:

من الواضح أنه ينبغي اختيار هذه البنود أو إنشاؤها مع مراعاة هدف القياس المحدد. وينبغــي أن يوجه هــذه العملية الوصف الدقيق لما يهدف المقيــاس عمله بالضبط. تذكر أن جميع البنود التي تشــكل مقياساً متجانســاً ينبغي أن تعكس المتغير الكامن الذي تقوم عليــه. ويمكن اعتبار كل بند بمثابة اختبار، في حد ذاته، لقوة المتغير الكامن. لذلك، ينبغي أن يعكس محتوى كل بند في المقام الأول المفهوم البنائي محل الاهتمام. وستشــكل البنود المتعــددة اختباراً أكثر ثباتــاً من كل بند من البنود على حدة، ولكن يجب أن يظل كل منها حساساً للدرجة الحقيقية للمتغير الكامن.

مــن الناحيــة النظرية، يتــم اختيار مجموعة جيــدة من البنود بطريقة عشــوائية من مجموعــة كلية أو شــاملة universe of items من البنود المتعلقــة بالمفهوم البنائي محل الاهتمام. ويفترض أن تكون المجموعة الشــاملة من البنود كبيرة بشــكل لا نهائي، وهو ما يحــول إلى حد كبير دون أي أمل في التعرف عليه في الواقع واســتخراج البنود منها بشــكل عشــوائي. ومع ذلك، ينبغي أن يوضع هذا الوضع المثالي في الاعتبار. إذا كنت تكتب بنوداً جديدة، كما هو الحال في كثير من الأحيان، يجب أن تفكر بشكل إبداعي في المفهوم البنائي الذي تســعى إلى قياســه. ما هي الطرق الأخرى التي يمكن بها صياغة بند ما للوصول إلى المفهوم البنائي؟ وعلى الرغم من أنه لا ينبغي ألا تتجاوز هذه البنود حدود تعريف المفهوم البنــائي المحدد، فإنه ينبغي أن تســتنفد هذه الإمكانيات أنواع البنــود الواقعة داخل تلك الحدود. ويتم تحديد خصائص المقياس بواسطة البنود التي يتكون منها. وإذا كانت البنود تعكس أو تجسد بشكل ضعيف المفهوم الذي عملت على توضيحه فترة طويلة وشاقة، فإن المقياس لن يعكس بدقة جوهر المفهوم البنائي.

ومن المهم أيضًا أن يكون «الشيء» الذي تشــترك فيه البنــود مفهوماً بنائياً حقيقياً وليس مجــرد فئة. وتذكر مرة أخرى أن نماذجنا لوضع وتطويــر المقياس تعتبر البنود مظاهر علنية لمتغير كامن مشترك هو سببها. يتم تحديد الدرجات على البنود المتعلقة بمفهوم بنائي مشترك بالدرجــة الحقيقية لهذا المفهوم البنائي. ومع ذلك، كما هو مذكور في الفصل الأول، لمجرد أن البنود تتعلق بفئة category مشتركة، فإن ذلك لا يضمن أن لها نفس المتغير الكامن الأساسي.

وغالبًـا ما تحدد مصطلحات مثـل الاتجاهات أو العوائق التي تحول دون الامتثال أو أحداث الحيـاة فئـات المفاهيم البنائية بدلاً من المفاهيم البنائية نفسـها. مجموعة من البنود التي سـتكون في نهاية المطاف أسـاس المقياس الأحادي البعد ينبغي ألا تتقاسم فقط التركيز على الاتجاهات، على سبيل المثال، بل على اتجاهات محددة، مثل الاتجاهات نحو معاقبة متعاطي المخدرات. يمكن للشـخص افتراض تصور حول سمة من سمات شخص ما - متغير كامن، إذا صح التعبير - من شأنه أن "يسبب" ردود على البنود التي تتناول معاقبة متعاطي المخدرات. ومن التحديات الكبيرة أن نتصور أن إحدى السـمات أو الخصائص تفسر الاتجاه بشكل عام. وينطبق الشيء نفسـه على الأمثلة الأخرى المذكورة. عادةً ما تكون عوائق الامتثال من أنواع عديدة. قد يمثل كل نوع (على سـبيل المثال، الخوف من اكتشـاف الأعراض، والقلق بشـأن تكاليف العلاج، وتوقع الألم، والمسافة إلى مرافق (مؤسسات) العلاج، وتصورات المناعة) متغير كامـن. قـد يكون هناك حتى علاقات متبادلة بين بعض المتغـيرات الكامنة. ومع ذلك، فإن كلاً مـن هذه العوائق هي مفهوم بنائي منفصل. ومـن ثم، فإن مصطلح العوائق يصف فئة من المفاهيم البنائية بدلاً من مفهوم بنائي على حدة يرتبط بمتغير كامن واحد. ولا ينبغي أن يُتوقـع من البنود التي تقيس مختلف المفاهيـم البنائية التي تندرج ضمن نفس الفئة (مثل التصورات حول المناعة والمخاوف بشـأن تكاليف العـلاج) أن تختلف في طريقة عمل البنود عندما تكون مظهراً لمتغير كامن مشترك.

التكرار:

من المفارقات أو التناقضات أن التكرار Redundancy يُعد سمة أو خاصية جيدة وسيئة للبنود الموجودة في المقياس. ويسـتلزم حل هذه المفارقة التمييز بين سـمات أو خصائص البنـود التـي تعزز وتدعم المقياس من خـلال التكرار وتلك التي لا تفعـل ذلك. لأن هذا الموضوع غالبًا ما يكون مصدرًا للالتباس، فسـوف أناقش ذلك بشيء من التفصيل، وسأقوم بادئ ذي بدء بالقضية التي تؤيد التكرار.

في هـذه المرحلـة من عمليـة وضع وتطوير المقيـاس، من الأفضل أن يكون أكثر شـمولاً، وأن تكـون كل الأمـور الأخـرى متسـاوية. التكرار ليس بالأمـر السيء عند وضـع المقياس. في الواقع، فإن النماذج النظرية التي توجه جهودنا لوضع وتطوير المقاييس تعتمد على التكرار.

في مناقشـــة معادلة تنبؤ "ســـبيرمان براون Spearman-Brown" في الفصل الثالث، أشرت إلى أن الثبـــات يختلـــف كدالة في عدد البنود، وكل الأمور الأخرى تكون متســـاوية. ونحن نحاول تجســـيد الظاهرة محل الاهتمام من خلال وضع وتطوير مجموعة من البنود التي تكشف عن هذه الظاهرة بطرق مختلفة. وباســـتخدام بنود متعددة تبدو متكررة، فإن المحتوى المشـــترك بين البنود ســـيتم جمعه عبر البنود بينما سيتم إلغاء الخواص غير ذات الصلة. وبدون التكرار، سيكون ذلك مستحيلاً.

ومع ذلك، فإن كل أشكال التكرار أمراً ليس مرغوباً فيه. ويتعلق التكرار المفيد بالمفهوم البنـــائي، وليس بالجوانب العرضيـــة أو الثانوية للبنود. تأمل في البندين التاليين: النســـخة الأصلية هي «شيء مهم حقًا نجاح طفلي»، والنســـخة المعدلة، «الشيء المهم حقًا هو نجاح طفـــلي». لم يتم تغيير شيء أكثر من أداة النكـــرة «a» إلى أداة التعريف"ال the" في بند ما بالتأكيد ســـيعطيك تكرارًا فيما يتعلق بالمحتوى الهام للبنـــد، ولكن البنود الأصلية والمعدلة ســـتكون أيضًا تكراراً فيما يتعلق بالعديد من الأشياء التي تريد تغييرها، مثل بنيتها النحوية الأساســـية واختيار الكلـــمات. ومن ناحية أخرى، فإن بندين مثل "ســـأفعل أي شيء تقريبًا لضـــمان نجاح طفلي" و"لا توجد تضحية أكبر إذا كان ذلك يســـاعد طفلي على النجاح" قد تكـــون تكراراً بشـــكل مفيد لأنها تعبر عن فكـــرة مماثلة بطرق مختلفـــة إلى حد ما. فهي تكـــرار فيما يتعلق بالمتغير محل الاهتمام ولكنها ليســـت تكراراً فيما يتعلق ببنيتها النحوية ومفرداتهـــا العرضية (الثانوية). وعندما يتم تجنب التكرارات التي ليســـت ذات صلة، فإن التكرارات ذات الصلة سوف تسفر عن مجموعات بنود أكثر ثباتاً.

عـــلاوة على ذلك، وعلى الرغم مـــن أن التكرار في الأداة النهائيـــة قد يكون غير مرغوب فيه، فإن المســـألة تكون أقل أهمية خلال المراحل المبكرة من وضع البنود. وبناءً على ذلك، قـــد تكون هناك قيمة وفائدة لصيغتي أو نســـختي البند المختلفتـــين في كلمة واحدة فقط لإدراجهـــما وتضمينهما في اختبار البند الأوَّلي. من خـــلال عمل ذلك، يمكن التحقق مما إذا كانـــت إحدى النســـختين للبند متفوقة (أي أكثر جودة) على الأخـــرى، ومن ثم يمكن إدراج وتضمين البند المتفوق في النسخة النهائية للمقياس. وقد تم تقديم حجة ضد التكرار جزئيًا وهي: أن ســـمات البنود السطحية المكررة مثل المفردات العرضية أو الثانوية (أي غير ذات الصلة بالمفهوم البنائي) أو البنية النحوية لا تُعد ميزة. قد تؤدي أوجه التشـــابه في الصياغة

التي لا صلة لها بالمفهوم البنائي إلى تفاعل المجيبين بشــكل مماثل مع البنود بطريقة تسفر عن تقدير مبالغ فيه للثبات. على سبيل المثال، إذا كانت هناك عدة بنود تبدأ بعبارة مشتركة (على ســبيل المثال، «عندما أفكر في الأمر، ...»)، قد يؤدي مجرد التشارك في تلك العبارة إلى ارتباط تلك البنود ببعضها البعض بشــكل أقوى. ولا يميز مؤشر الثبات مثل «ألفاكرونباخ» بين التغاير للبنود الناشئ عن تلك الصياغة المشتركة والتغاير الذي يعزى إلى التأثير المشترك للمتغير محل الاهتمام. ومن ثم، سيكون تقدير الثبات مبالغاً فيه أو متضخماً.

وفي حين أن الســمات النحوية المماثلة أو غيرها من الســمات الســطحية الأخرى يمكن أن تشــكل تشــابهاً غير مرغوب فيه في المحتوى، قد يؤدي التكرار غير المرتبط بشكل كامل بالمفهوم البنائي موضع الاهتمام إلى حدوث مشكلة في بعض الظروف. ويمكن أن يحدث هذا عندمــا تختلف بعض البنود عن معظم البنود الأخرى في مجموعة فيما يتعلق بالخصوصية أو التحديد. على ســبيل المثال، لنفكر مليــاً في مقياس افتراضي يهدف إلى قياس الاتجاهات نحــو محبي الحيوانات الأليفة. قد يكون من المناســب إدراج مجموعة متنوعة من البنود في المقياس. وقد تكون هناك بنود أخرى أيضاً، وإن كانت ذات صلة بالمفهوم البنائي موضع الاهتــمام، محــددة أكثر من اللازم - وبالتالي مكررة إلى حد كبير - للعمل بشــكل جيد. قد تكــون البنود التالية «عشــاق الببغاء الرمادي الأفريقي هم أشــخاص لطاف» و«أعتقد أن الناس الذين يحبون الببغاوات الرمادية الأفريقية هم أشــخاص طيبون» متشــابهة إلى حد كبير ليس فقط بســبب تشابه البنية النحوية ولكن بسبب احتمال أن يكون المحتوى الذي يتقاســمه البندين ملائماً ولكنه محدد بشــكل مبالغ فيه. قد يقومون بســحب البند الذي تم تحديده ككل بعيداً عن المتغير الكامن المســتهدف أو المقصود (الاتجاهات نحو محبي الحيوانــات الأليفة) إلى بديل أكثر تحديداً للمتغير الكامن (الاتجاهات نحو عشــاق الببغاء الرمــادي الأفريقــي). ومع الأخذ في الاعتبار وجود مجموعة كبــيرة من الحيوانات الأليفة، هناك بندان حول نوع معين وغير مألوف من الحيوانات الأليفة متشــابهان بشــكل كبير، ومن المحتمل أن يقوضا الغرض من الأداة.

وبصــورة أعم، فإن مدى عموميــة أو تحديد المفهوم البنائي موضــع الاهتمام يمكن أن يغــير من تأثير التكرار. وعلى الرغم من أن مثال الببغاء الرمادي الأفريقي قد يبدو متطرفاً بعــض الــشيء، فإن إدراج بنود لا تتطابــق مع خصوصية أو تحديد المفهــوم البنائي موضع

الاهتمام يمكن أن يحدث في سياق أقل غرابة. على سبيل المثال، في أداة تم تصميمها لتعكس أو تجسد جميع جوانب العاطفة، فإن العديد من البنود المتعلقة بالقلق قد تمثل مشكلة. ومن المحتمل أن تكون الارتباطات بين هذه البنود أكبر من الارتباطات بين تلك البنود وغيرها من البنود التي لا تتعلق بالقلق. وكنتيجة لذلك، قد تشكل هذه البنود مجموعة فرعية من بنود القلق داخل المجموعة الأوسع من بنود العاطفة. وهذا يمكن أن يسبب عدداً من المشكلات. أولاً، قد يؤدي ذلك إلى تقويض أو الإضرار بأحادية البعد لمجموعة البنود (التي ستكون مشكلة إذا كانت نية أو هدف الباحث هو وضع مقياس واحد لمتغير أحادي البعد). أيضاً، قد يؤدي ذلك إلى إنشاء نقطة اتصال غير مقصودة ينتج عنها بنود أكثر تشابهاً لتلك الموجودة في عنقود (مجموعة) القلق التي يبدو أنها أفضل أداءً من تلك البنود التي تكون أقل تشابهاً. فعلى سبيل المثال، على الرغم من أن بنداً متعلقاً بالقلق وآخر متعلق بالخوف قد يكونان على نفس القدر من الأهمية بالنسبة لوجهة النظر الواسعة للعواطف، فإن البند الأول قد يسهم بقوة أكبر في الثبات من البند الأخير إذا كانت هناك أغلبية عددية لبنود القلق في الأداة. ونتيجة لذلك، قد يتجاوز متوسط ارتباط بند القلق متوسط ارتباط بند الخوف، مما يؤدي إلى زيادة مساهمته في تقدير الثبات. والواقع أن الأداة التي تتضمن تمثيلاً مبالغاً فيه لبنود القلق على النحو الموصوف لن تعبر عن جميع العواطف بالتساوي ولكن ستكون منحازة نحو القلق.

وعلى النقيض من ذلك، فإن نفس النوع من بنود القلق التي توصف بأنها تمثل إشكالية في الفقرة السابقة قد لا تكون تكراراً بشكل مفرط أو مبالغ فيه في أداة ذات تركيز أضيق. ومن الواضح أنه إذا كانت الأداة مصممة لتقييم القلق، فينبغي أن تكون جميع البنود ذات صلة بذلك المتغير وأن التشابه لن يكون حالة من حالات التكرار غير المرغوب فيه. وعلى النقيض من ذلك، فإن البنود التي تتضمن عبارة أكثر عمومية، مثل «مشاعري العامة»، قد تشكل مجموعة فرعية إذا تم تضمينها في حالة مقياس القلق بسبب تركيزها العاطفي غير المحدد. غير أن ما يبدو على أنه مشكلة تكرار قد يكون في الواقع مسألة مدى تطابق البنود مع خصوصية وتحديد المفهوم البنائي الذي يعتزم الباحث تقييمه.

في أداة تهدف إلى تجسيد متغير أكثر تحديدًا، فمن المرجح أن البنود سوف تظهر أو تبدو أكثر تشابهًا مع بعضها البعض. عادةً، على سبيل المثال، ستبدو البنود الموجودة في مقياس

يقيــس القلــق في مجال الخطابة العامة أكثر تشــابهاً مع بعضها البعــض (نظرًا لخصوصية المتغير موضع الاهتمام) من البنود الموجودة في مقياس يقيس الحالات العاطفية على نطاق أوسع. وقد لا تكون هذه مشكلة طالما كانت أوجه التشابه مرتبطة بالمفهوم البنائي موضع الاهتمام. وكما ذكرنا من قبل، فإن البنود المتشــابهة بقدر ارتباطها بمتغيرها المقصود وليس في أي مجال آخر يمكن أن تكون بنوداً جيدة.

عدد البنود:

من المستحيل تحديد عدد البنود التي ينبغي إدراجها أو تضمينها في تجمع أوَّلي. ويكفي القول إنك تريد أكثر بكثير مما تخطط لإدراجه في المقياس النهائي. تذكر أن ثبات الاتســاق الداخــلي هو دالة لمدى قوة ارتباط البنود مع بعضهــا البعض (وبالتالي مع المتغير الكامن) وعــدد البنود الموجودة لديــك في المقياس. وبما أن طبيعة الارتباطــات المتبادلة بين البنود تكــون غير معروفة عادةً في هذه المرحلــة من وضع وتطوير المقياس، فإن وجود الكثير من البنود يُعد شــكلاً من أشــكال التأمين ضد ضعف الاتســاق الداخلي. كلما زاد عدد البنود التي لديك في المجموعة المتوفرة لديك في تجمع البنود، كان الأمر أكثر صعوبة بشأن اختيار البنود التي ســتؤدي المهمة التــي تنوي القيام بها. وليس من غــير المعتاد البدء بمجموعة مــن البنود التي يبلغ حجمها ثلاثــة أو أربعة أضعاف الحجم النهائي للمقياس. وهكذا، قد يتطــور مقيــاس مكون من عشرة بنود من مجموعة تضم ٤٠ بنــداً. وإذا كان من الصعب بشــكل خاص توليد وإنشاء بنود في مجال معين من مجالات المحتوى أو إذا كانت البيانات التجريبية تشــير إلى أنه لا توجد حاجة إلى بنود عديدة لتحقيق اتســاق داخلي جيد، فقد يكون التجمع الأوَّلي للبنود أصغر بنسبة ٥٠٪ من حجم المقياس النهائي.

بشــكل عــام، كلما كان تجمع البنود أكبر، كان ذلك أفضــل. ومع ذلك، فمن المؤكد أنه من الممكن تطوير تجمع كبير جداً من البنود بحيث لا يمكن إدارته في مناســبة واحدة لأي مجموعة واحدة من الأشــخاص. إذا كان تجمع البنود كبيراً بشكل استثنائي، يمكن للباحث أن يلغي ويحذف بعض البنود وفقاً لمعايير مســبقة، مثل عدم الوضوح، والأهمية أو الصلة المشكوك فيها، أو التشابه غير المرغوب فيه مع بنود أخرى.

بدء عملية كتابة البنود:

غالبًـا مـا يكون البدء في كتابــة البنود هو أصعب جزء في عملية توليــد وتكوين البنود. واســمحوا لي أن أصف كيف أبدأ هذه العملية. في هذه المرحلة، أكون أقل اهتماماً بجودة البنود ونوعيتها من مجرد التعبير عن الأفكار ذات الصلة. وأبدأ في كثير من الأحيان بعبارة أو فقرة تمثل إعادة صياغة للمفهوم البنائي الذي أريد قياسه. على سبيل المثال، إذا كنت مهتمًا بوضع وتطوير مقياس للإدراك الذاتي لتأثير الإعلانات التجارية، فيمكن أن أبدأ بالــعبارة أو الفقرة التالية «أنا سريع التأثر بالإعلانات التجارية». بعد ذلك سأحاول توليد وإنشاء فقرات وعبارات إضافية تدور حول نفس الفكرة بشكل مختلف إلى حد ما. قد تكون العبارة التالية «الإعلانات التجارية تؤثر عليّ كثيراً». وسأستمر بهذه الطريقة أو الأسلوب، مع عدم الاهتمام بوجــود أي معايير للجودة تقريبــاً في العبارات. وهدفي في هذه المرحلــة المبكرة هو مجرد تحديــد مجموعة كبيرة من الطرق التي يمكن من خلالها توضيح وتحديد المفهوم الأســاسي الذي تســتهدفه الأداة. وأثناء الكتابة، فإنني قــد أبحث عن طرق بديلة للتعبير عن الأفكار النقدية. على سبيل المثال، قد أستبدل عبارة «الإعلانات التجارية» بعبارة هي «الأشياء التي أراها في إعلانات التلفزيون أو المجلات» في المجموعة التالية من الجمل. وأنا أجد أن الكتابة بسرعة ودون نقد يُعد أمراً مفيداً. بعد توليد وإنشاء ربما ثلاثة أو أربعة أضعاف عدد البنود التــي أتوقع إدراجها وتضمينها في الأداة النهائية، ســوف ألقى نظــرة على ما قمت بكتابته. والآن قد حان الوقت الحاســم أو المهم. حيث يمكن فحص البنود للوقوف على مدى جودة تجســيدها للأفكار الأساســية ومدى وضوح الصياغة. وتحدد الأقسام التالية بعض خصائص البنود المحددة لتجنبها أو إدراجها في عملية الاختيار من قائمة العبارات الأصلية وتنقيحها.

خصائص البنود الجيدة والبنود السيئة:

إن سرد جميع الأشياء والأمور التي تجعل بنداً ما جيداً أو سيئاً تعد مهمة مستحيلة. ومن الواضح أن مجال أو نطاق المحتوى له تأثير كبير محتمل على جودة البند. ومع ذلك، هناك بعــض الخصائص التي تفرق وتميز على نحو موثوق أفضل البنود عن أســوأ البنود. وتتعلق معظم هذه الخصائص بالوضوح. وكما هو مبين في الفصل الأول، ينبغي أن يكون البند الجيد واضحاً لا لبس فيه. وينبغي التخلي عن الأسئلة التي تضع المجيب في مأزق وارتباك.

يجب على واضعي أو مطوري المقياس تجنب البنود الطويلة بشــكل اســتثنائي، حيث إن طــول البنــد عادةً ما يزيد من التعقيد ويقلل من الوضوح. غير أنه ليس من المرغوب فيه أو المســتصوب التضحية بمعنى أو بمدلول بند ما لصالح الإيجــاز. إذا كان تعديل البند ضروريًا لإيصال ونقل الغرض من أحد البنود، فقم بإدراجه. ومع ذلك، تجنب الإسهاب والإطناب غير الضروري. وبوجهٍ عام، سيكون بند ما مثل «كثيراً ما أجد صعوبة في التعبير عن وجهة نظري» أفضل من عبارة أطول دون داعٍ، مثل «من الإنصاف القول إن أحد الأشياء التي يبدو لي أنني أواجه فيها مشكلة في كثير من الوقت هو إيصال ونقل وجهة نظري إلى أشخاص آخرين».

وثمــة اعتبار آخر ذي صلة باختيــار أو تطوير البنود هو مســتوى صعوبة القراءة الذي تتم به كتابة البنود. هناك مجموعة متنوعة من الأســاليب (على ســبيل المثال، دايل وشل، Dale & Chall, 1948؛ فري، Fry, 1977) لتعيين مستويات الصف بالنسبة للكلام النثري بما في ذلك بنود المقياس. عادةً ما تساوي هذه الكلمات والجمل الأطول مع مستويات القراءة الأعلى. ويفترض أن قراءة معظم الصحف المحلية تتطلب مستوى قراءة للصف السادس.

يحدد "فري" Fry (1977) عدة خطوات لتحديد مســتوى القراءة كمياً. الخطوة الأولى هــي تحديد عينة من النص الذي يبــدأ بالكلمة الأولى من الجملة ويحتوي على ١٠٠ كلمة بالضبط. (بالنســبة للمقاييس التي تحتوي على عــدد قليل من البنود فقط، قد تضطر إلى تحديد جزء مناســب من ١٠٠ وتأســيس الخطوات اللاحقة على هذه النســبة.) بعد ذلك، احســب عدد الجمــل الكاملة والمقاطع اللفظيــة الفردية في عينة النص. ويتم اســتخدام هذه القيم كنقاط إدخال لمخطط أو رســم بياني يوضح مســتوى الصف المكافئ أو المعادل لتوليفــات وتركيبــات مختلفة من الجملة وتعداد المقاطع اللفظيــة من العينة المكونة من ١٠٠ كلمة. يشــير الرسم البياني إلى أن متوسط عدد الكلمات والمقاطع لكل جملة لمستوى القــراءة في الصــف الخامس هــو ١٤ و١٨ على التوالي. وتبلغ الجملة المتوســطة في الصف الســادس ١٥ أو ١٦ كلمة وما مجموعه ٢٠ مقطعاً لفظياً؛ أما جملة مســتوى الصف السابع فتتكون من حوالي ١٨ كلمة و٢٤مقطعاً لفظياً. ويمكن أن تسفر الجمل الأقصر التي تحتوي على نســبة أعلى من الكلمات الأطول أو الجمل الأطول مع عدد أقل من الكلمات الطويلة عن مستوى يعادل الصف. على سبيل المثال، يتم تصنيف جملة مكونة من ٩ كلمات و١٣ مقطعاً لفظياً (أي ما يصل إلى ٤٤٪ كلمات متعددة المقاطع او متشــابكة) أو جملة مكونة

مـــن ١٩ كلمة و٢٢ مقطعاً لفظياً (أي، ما لا يزيد عن ١٤٪ من الكلمات متعددة المقاطع أو متشابكة) على حد سواء على أنها تعادل مستوى القراءة للصف السادس. ومن المحتمل أن يكون الهدف من تحقيق مستوى للقراءة بين الصفين الخامس والسابع هدفاً مناسباً لمعظم الأدوات التي سيتم استخدامها مع المجتمع الإحصائي العام. فعلى سبيل المثال، كُتبت بنود مقاييس موضع الضبط أو السيطرة الصحية المتعدد الأبعاد على مستوى القراءة من الصف الخامس إلى الصف الســابع. وهناك بند نموذجي في هذا المســتوى من القراءة هو «معظم الأشــياء التي تؤثر على صحتي تحدث لي عن طريــق الصدفة» (Wallston et al., 1978). هــذا البند يتكــون من ١١ كلمة و١٥مقطعاً لفظياً مما يضع هذا البند في مســتوى القراءة للصف السادس.

ويشــير فــري (1977) Fry إلى أنه ينبغي النظر في العوامــل الدلالية والنحوية وأخذها في الاعتبــار عنــد تقييم صعوبة القراءة. ونظرًا لأن الكلمات القصيرة تميل إلى أن تكون أكثر شــيوعًا، وتميل الجمل القصيرة إلى أن تكون أبسط من الناحية النحوية، فإن الإجراء الخاص بـ "فري" يعد بديلاً مقبولًا لطرق تقييم الصعوبة الأكثر تعقيداً. ومع ذلك، وكما هو الحال مع المعايير الأخرى للكتابة أو اختيار البنود الجيدة، يجب على الشــخص اســتخدام المنطق السليم في تطبيق أساليب مستوى القراءة. بعض العبارات المختصرة التي تحتوي فقط على كلمات قصيرة ليســت أوَّلية أو بســيطة. على ســبيل المثال، عبارة مثل «تجنب الازدراء في الكاســك» Eschew casque scorn من المرجح أن تربك شــخصاً ما بمستوى تعليم ابتدائي أكثر من عبارة "ارتدِ خوذتك" Wear your helmet، على الرغم من أن كل جملة يوجد بها ثلاث كلــمات وأربعة مقاطع لفظية. وهناك مصدر آخر للارتباك والالتباس المحتمل يجب تجنبه وهو كلمات النفي «الســلبيات» المتعددة. العبارة التالية «أنا لا أؤيد قيام الشركات بوقف تمويل الجماعات المناهضة للأســلحة النووية» تُعد أمراً مربكاً أكثر بكثير من العبارة التالية «أنا أؤيد الدعم الخاص المستمر للمجموعات التي تدعو إلى حظر الأسلحة النووية». (مــن المفيد أيضًا ملاحظــة أن هاتين العبارتين قد تعبران عن مواقف مختلفة بشــأن هذه القضية. فعلى سبيل المثال، قد تنطوي هذه العبارة الأخيرة على تفضيل دعم القطاع الخاص عن دعم القطاع العام للجماعات المعنية).

وينبغي أيضاً تجنب ما يسمى بالبنود المزدوجة double-barreled items. وهي البنود التي تنقل فكرتين أو أكثر بحيث أن الإجابة على البند بالموافقة قد يشـير إلى أي من الأفكار أو كليهما. العبارة التالية «أنا أؤيد الحقوق المدنية لأن التمييز وعدم المساواة جريمة ضد الله» هي مثال على البند المزدوج. إذا كان الشخص يدعم الحقوق المدنية لأسباب أخرى غير الإساءة إلى الذات الإلهية (على سبيل المثال، لأنها جريمة ضد الإنسانية)، فكيف ينبغي أن يجيب/ تجيب على البند؟ فقد تعبر الإجابة السلبية بشكل غير صحيح عن عدم تأييد الحقوق المدنية، والإجابة الإيجابية قد تنسب بشكل غير صحيح دافع إلى دعم المجيب.

وثمة مشكلة أخرى يجب على واضعي ومطوري المقياس تجنبها هي إشارات الضمير الشخصي pronoun الغامضة. البند التالي "يجب على القتلة ومرتكبي جرائم الاغتصاب ألا يلتمسوا العفو من السياسيين لأنهم حثالة الأرض" قد يعبر عن مشاعر بعض الناس بغض النظر عن إشارة الضمير. (ومع ذلك، عادةً ما ينوي مطور مقياس أن يكون أكثر وضوحاً حول ما يعنيه البند). هذه الجملة يجب أن تكون سيئة أو غير مرغوبة مرتين. فبالإضافة إلى إشارة الضمير الغامضة، فهي بند مزدوج. وتؤدي المُعدِّلات modifiers التي تكون في غير محلها إلى حالات غموض والتباس مماثلة لإشارات الضمير الغامضة: الجملة التالية: "ينبغي لأعضائنا في الكونغرس أن يعملوا بجد لإضفاء الشرعية (لتقنين البغاء) على البغاء في مجلس النواب" هي مثال على هذه المُعدِّلات. استخدام أشكال أو نماذج الصفات بدلاً من أشكال الأسماء يمكن أيضاً أن يحدث ارتباكاً وتشويشاً غير مقصود. النظر في الاختلافات في المعنى بين All vagrants should be given a schizophrenic assessment "ينبغي إعطاء جميع المتشردين تقييم المصاب بانفصام الشخصية" وAll vagrants should be given a schizophrenia assessment "ينبغي إعطاء جميع المتشردين تقييم انفصام الشخصية".

إن الكلمات الفردية ليست هي المصادر الوحيدة لغموض البند. فالجملة بأكملها يمكن أن يكون لها أكثر من معنى واحد. ولم أرَ في الواقع سوى دراسة استقصائية واحدة للسلوك الجنسي للمراهقين تضمنت بنداً لتقييم تعليم الوالدين. وبالنظر إلى سياق الدراسة الاستقصائية ككل، فإن الصياغة كانت مؤسفة: «إلى أي مدى ذهبت أمك إلى المدرسة؟» (How far did your mother go in school?) وقد فشل الباحثون تماماً في التعرف على المعنى غير المقصود لهذه الجملة إلى الحد الذي أثار سخط مجموعة من المهنيين أثناء عرض

حلقــة النقاش. وأظن أن عدداً لا بــأس به من المجيبين المراهقين قد ضحكوا أيضاً من هذا البند. ولا يُعرف كيف أثر ذلك على ردودهم على ما تبقى من الاستبانة.

البنود ذات الصياغة الإيجابية والبنود ذات الصياغة السلبية:

يختــار العديــد من مطــوري المقاييس كتابة البنــود ذات الصياغة الســلبية التي تمثل مســتويات منخفضــة أو حتى غياب المفهوم البنــائي موضع الاهتمام وكذلــك البنود ذات الصياغة الإيجابية الأكثر شيوعًا، والتي تمثل وجود المفهوم البنائي موضع الاهتمام. والهدف مــن ذلــك هو الوصول إلى مجموعة من البنود، يشــير بعضها إلى مســتوى عالٍ من المتغير الكامن عند الموافقة عليه وإقراره، وأخرى تشــير إلى مســتوى عــالٍ عند عدم الموافقة على المتغير الكامن. على ســبيل المثال، يتضمن مقياس «روزنبرغ» Rosenberg (1965) لتقدير الذات، على ســبيل المثال، بنوداً تشــير إلى التقدير العالي للذات (على ســبيل المثال، "أشعر بأن لديّ عددًا من الصفات الجيدة") وأخرى تشــير الى تقدير للذات منخفض (على ســبيل المثال، "أشــعر بالتأكيد أنني عديم الفائدة في بعض الاحيان"). والغرض من صياغة البنود بشكل إيجابي أو سلبي على حد سواء ضمن نفس المقياس هو عادةً تجنب القبول أو التأكيد أو التحيز في الموافقة. تشــير هذه المصطلحات التي تســتخدم بالتبادل إلى ميل المجيب إلى الموافقة على البنود بغض النظر عن محتواها ومضمونها. إذا كان المقياس، على سبيل المثال، يتكون من بنود تعبر عن درجة عالية من تقدير الذات، فإن تحيز القبول قد يؤدي إلى نمط من الاســتجابات التي تبدو وكأنها تشــير إلى تقدير للذات مرتفع للغاية. وإذا كان المقياس مكونًا من عدد متساوٍ من البنود ذات الصياغة الإيجابية والسلبية، فمن ناحية أخرى، يمكن التمييز بين تحيز الموافقة والدرجة المتطرفة من تقدير الذات عن بعضهما البعض عن طريق نمط الاســتجابات. من شــأن "الموافقة" أن يؤيد البنود التي تشير إلى تقدير الذات العالية والمنخفضة على حد ســواء، في حين أن الشــخص الذي كان يحظى بتقدير كبير حقًا سيؤيد بشدة البنود التي تحظى بتقدير كبير ويؤيد بصورة سلبية البنود ذات التقدير المنخفض.

ولســوء الحــظ، قد يكون هنــاك ثمن يجب دفعــه مقابل إدراج وتضمــين البنود ذات الصياغة الإيجابية والســلبية. قد يكون تغيير الاتجاه (تحويل الصياغة الســالبة إلى موجبة والعكس) في قطبية العناصر مربكة للمستجيبين، خاصة عند ملء استبيان طويل. وفي مثل

هــذه الحالة، قد يصبح المســتجيبون في حيرة من أمرهم حول الفــرق بين التعبير عن قوة اتفاقهــم مع جملة (بند) مــا، بغض النظر عن قطبيتها، وبين التعبير عن قوة الســمة التي يتم قياســها (التقدير، على ســبيل المثال). وبصفتي باحثًا في العلوم الاجتماعية التطبيقية، فقــد رأيت العديد من الأمثلة للبنود التي صيغــت في الاتجاه المعاكس تؤدي إلى أداء سيء للبنود. على سبيل المثال، وصف «ديفيليس وكالاهان» DeVellis and Callahan (1993) بديــلاً أقصر وأكثر تركيزاً لمؤشر الاتجاهات نحو مرض الروماتيزم (وهو اســم مؤســف، لأن الأداة لا تقيّم الاتجاه ولا تُعد مؤشرًا). لقد اخترنا بنوداً من النســخة الأصلية والأطول بناءً عــلى معايــير تجريبية وانتهى بنا الأمــر إلى أربعة بنود تعبر عن ردود الفعل الســلبية على المــرض وأحدهــم يعبر عن القدرة عــلى مواجهة (التغلب على) المرض بشــكل جيد. وكان القصــد من ذلك هو أن يقوم المســتخدمون بعكس درجة تســجيل بند "المواجهة" بحيث تكون جميع البنود معبرة عن الشــعور بالعجز. وفي الآونة الأخيرة، درس «كوري ،كالاهان، وديفيليــس» Currey, Callahan, and DeVellis (2002) أداء هذا العنصر الفردي الذي صيغ في الاتجاه الإيجابي. وقد كان أداؤه ضعيفاً (منخفضاً) باســتمرار. وعندما تمت إعادة صياغة البند ببســاطة عن طريق إضافة كلمة (لا) لتغيير تكافؤه بحيث يكون متســقاً مع البنود الأخرى، فقد تحســن أداؤه بشــكل كبير. ونعتقد أنه على الرغم من إدراك العديد مــن المجيبين للتكافؤ المختلف للبند الأصلي، إلا أن آخرين لم يدركوا ذلك. ومن شــأن ذلك أن يؤدي إلى وجود نســبة من الأفراد الذين أكــدوا الارتباطات الإيجابية مع البنود الأربعة الأخرى ونســبة أخرى لهم نفس الارتباطات الســلبية. ونتيجة لذلك، بالنسبة للعينة ككل، فــإن الارتباطات بين هذا البند والبنود الأربعة الأخرى ســوف تتضاءل أو تنخفض بشــكل ملحوظ، وبالتالي، ســوف ينتج نوع الأداء غير المرضي الذي لاحظناه بالنســبة للبند الأصلي عكــس بنــد التكافؤ. توحــي لي التجربة أو الخبرة الشــخصية مع العينــات المعتمدة على المجتمع المحلي أن عيوب ومساوئ البنود التي صيغت في اتجاه معاكس تفوق أي فوائد.

الخلاصة:

يجب أن يكون تجمع pool البنود مصدراً غنيًا يمكن أن ينشــأ منه مقياس. وينبغي أن يتضمــن عدداً كبيراً من البنود ذات الصلة بالمحتوى موضــع الاهتمام. التكرار فيما يتعلق

بالمحتوى هو أصل، وليس التزام. وهو أســاس ثبات الاتســاق الداخــلي الذي هو، بدوره، أساس الصدق. وينبغي ألا تنطوي البنود على "صفقة شاملة أو مجموعة مقاييس" تجعل من المستحيل على المجيبين تأييد جزء من البند دون تأييد جزء آخر قد لا يكون متسقاً مع الجزء الأول. سواء تم تضمين البنود ذات الصياغة الإيجابية والسلبية في تجمع البنود أم لا، فإن صياغتها يجب أن تتبع القواعد اللغوية المعمول بها. وســوف يساعد ذلك على تجنب بعض مصادر الغموض التي نوقشت أعلاه.

الخطوة الثالثة: تحديد شكل أو تنسيق القياس:

توجد أشكال وصيغ عديدة للأسئلة. يجب على الباحث أن يفكر مبكراً في الشكل الذي ســيكون عليه المقياس. وينبغي أن تحدث هذه الخطوة بشــكل متزامن مع عملية توليد وإنشــاء البنود بحيث يتم التوافق بين الاثنين. على ســبيل المثال، قد يكون إنشاء وتكوين قائمة طويلة من العبارات التوضيحية مضيعة للوقت إذا كان شــكل الاســتجابة الذي وقع عليــه الاختيار في نهاية المطاف عبارة عن قائمة مرجعيــة مؤلفة من بنود مكونة من كلمة واحدة. علاوة على ذلك، النماذج النظرية المقدمة في وقت سابق هي أكثر اتساقاً مع بعض صيغ أو أشكال الاستجابة من غيرها. وبصفة عامة، فإن المقاييس التي تتكون من بنود قابلة للتســجيل على مقياس متصل ويتم تجميعها لتشــكل درجة المقياس هي الأكثر توافقاً مع التوجه النظري الوارد في هذا الكتاب. ومع ذلك، ســأناقش في هذا القسم الأشكال والصيغ الشــائعة التي تختلف عن النمــط الذي تتضمنه النماذج النظرية التي تمت مناقشــتها في الفصل الثاني وكذلك الأشكال التي تتقيد بهذا النمط.

مقياس ثرستون (Thurstone):

هناك عدد من الإســتراتيجيات العامة لبناء المقاييس تؤثر على شــكل البنود وخيارات الاســتجابة. إحدى هذه الإســتراتيجيات هي مقياس «ثورستون»Thurstone. قد يساعد التشــبيه على توضيح كيفية عمل مقياس "ثورستون". تم تصميم شوكة رنانة لتهتز بتردد معــين. إذا قمت بضربها، فإنها ســتهتز في هذا التردد وســتؤدي إلى إصــدار نغمة معينة. وعــلى العكــس من ذلك، إذا وضعت الشــوكة بالقرب من مصدر نغمــة ينتج نفس تردد

شـوكة الرنين، تبدأ الشـوكة بالاهتزاز. ﺑﻤعنى أن الشـوكة الرنانة عبارة عن "جهاز كشف تـردد"، تهتز في وجـود موجات صوتية من التردد الرنان وتظل بـدون حركة في وجود كل الترددات الأخرى. تخيل سلسـلة من الشوكات الرنانة المرتبة في مصفوفة ما بحيث عندما يتحـرك إحداها من اليسـار إلى اليمين على طول المصفوفة، تتوافق الشـوكات الرنانة مع أصـوات الـتردد الأعلى تدريجيًا. ضمن نطاق تردد الشـوكة الرنانة، يمكن اسـتخدام هذا الترتيـب لتحديد تردد النغمة. وبعبارة أخرى، يمكنك تحديد تردد النغمة من خلال رؤية أي شـوكة تهتز عند تشـغيل النغمة. إن مقياس «ثورسـتون» تم إعـداده ليعمل بنفس الطريقـة. يحاول واضع ومطور المقياس توليد وإنشـاء بنود تسـتجيب بشـكل متفاوت لمستويات معينة من السمة المعنية. عندما تتطابق «درجة» بند معين مع مستوى السمة التـي يمتلكهـا المجيب، فإن هذا البند سيشـير إلى هـذا التطابق. وفي كثـير من الأحيان، تتألف الإشارة من اسـتجابة إيجابية للبنود التي يتم «ضبطها» على المستوى المناسب من السـمة واستجابة سـلبية لجميع البنود الأخرى. وعادةً ما يتم تحديد عملية الضبط (أي تحديد مستوى المفهوم البنائي الذي يستجيب له كل بند) عن طريق قيام المحكمين بوضع مجموعة كبيرة من البنود في مجموعات تقابل بفواصل أو فترات زمنية متساوية من حيث أهمية المفهوم البنائي أو قوته.

هذه فكرة ممتازة للغاية. ويمكن تطوير البنود لتتوافق مع درجات شـدة مختلفة من السـمة، ويمكن أن تكون متباعدة بحيث تمثل فواصل أو فترات متسـاوية، ويمكن تنسيقها وصياغتها باسـتخدام خيارات الاسـتجابة موافـق وغير موافق، على سـبيل المثال. ويمكن للباحـث إعطاء هـذه البنود للمجيبين ثم يفحـص إجاباتهم لمعرفة البنـود التي أدت الى الاتفاق. ونظراً لأن البنود ستتم معايرتها مسبقًا فيما يتعلق بحساسيتها لمستويات محددة من الظاهرة، فإن الاتفاقات سـتحدد مقدار السـمة التي يمتلكها المجيب. قد يؤدي اختيار البنـود التي تمثل فترات متسـاوية عبر البنود إلى خصائص قياس مرغـوب فيها للغاية لأن الدرجات ستكون قابلة للإجراءات الرياضية القائمة على القياس الفتري أو المسافي.

قد يبدو جزء من مقياس «ثورستون» الافتراضي لقياس تطلعات الآباء للتحصيل التعليمي والوظيفي لأطفالهم كما يلي:

- تحقيق النجاح هو السبيل الوحيد لطفلي أن يسدد الجهود التي بذلتها كوالد.	موافق	غير موافق ....
- الذهــاب إلى كلية جيدة والحصول على عمل جيد أمرٌ مهمٌ ولكن ليس ضرورياً لسعادة طفلي.	موافق	غير موافق ....
- الســعادة ليســت لها علاقة بتحقيق الأهداف التعليمية أو المادية.	موافق	غير موافق ....
- زخــارف النجاح ذات القيمــة المعتادة ليســت عائقاً أمام السعادة الحقيقية.	موافق	غير موافق ....

وكما يشــير نونالي (1978) Nunnally، فإن تطوير مقياس ثورســتون حقيقي وصحيح أصعــب بكثير من وصفه. وأنه من الصعب للغاية العثور على البنود التي «يتردد صداها» باســتمرار على مســتويات محددة من هذه الظاهــرة. وغالباً ما تفوق المشــاكل العملية المرتبطة بهذه الطريقة مزاياها ما لم يكن لدى الباحث ســبب مقنع لرغبته في نوع المعايرة التــي يوفرهــا. على الرغم مــن أن مقياس «ثرســتون» هو طريقة أو نهــج مثير للاهتمام ومناسب في بعض الأحيان، إلا أنه لن تتم الإشارة إليه في بقية هذا الكتاب. ومع ذلك، لاحظ أن الأساليب القائمة على نظرية استجابة البند، التي تمت مناقشتها في فصل لاحق، تشترك في العديد من أهداف مقياس «ثورستون» مع اتباع نهج مختلف إلى حد ما لتحقيقها.

مقياس جتمان (Guttman):

مقياس «جتمان» Guttman هو سلســلة من البنود التي تؤثر تدريجياً على مستويات أعــلى من الســمة. وهكــذا، ينبغي للمجيب أن يؤيــد ويوافق على مجموعــة من البنود المتجاورة إلى أن يتجاوز، عند نقطة حرجة، مقدار الســمة التي تســتخدمها البنود ويمتلكها الشــخص. ولا يجب تأييد وإقرار أي من البنود المتبقية. وتتوافق بعض البيانات الوصفية البحتة مع مقياس "جتمان". على ســبيل المثال، قد تطرح سلسلة من أسئلة المقابلة، "هل تدخن؟" "هل تدخن أكثر من ١٠ ســجائر في اليوم؟" "هل تدخن أكثر من علبة ســجائر في اليوم؟" وهكذا. كما هو الحال في هذا المثال، فإن إقرار أي بند محدد على مقياس "جتمان"

يعني ضمنياً تأكيد جميع البنود السـابقة. ويشـير أعلى بند يعطي رداً إيجابياً إلى مستوى السـمة الذي يمتلكه المجيب. لاحظ أنه، في حين أن كلاً من مقياس «ثرستون»Thurstone ومقياس "جتمان" Guttmanيتكونان من بنود متدرجة، فإن التركيز ينصب على اسـتجابة إيجابيـة واحدة في الحالة الأولى، ولكـن في الحالة الأخيرة ينصب التركيز على نقطة الانتقال أو التحول من الاسـتجابات الإيجابية إلى السلبية في الحالة الأخيرة. قد تبدو نسخة مقياس "جتمان" من مقياس الطموح الأبوي السابق كما يلي:

- تحقيق النجاح هو السـبيل الوحيد لطفلي لسداد جهودي كوالد.	موافق	غير موافق ....
- الذهـاب إلى كلية جيدة والحصول على عمل جيد أمرٌ مهمٌ ولكن ليس ضرورياً لسعادة طفلي.	موافق	غير موافق ....
- السـعادة ليسـت لها علاقة بتحقيق الأهداف التعليمية أو المادية.	موافق	غير موافق ....
- زخـارف النجاح ذات القيمـة المعتادة ليسـت عائقاً أمام السعادة الحقيقية.	موافق	غير موافق ....

يمكن أن تعمل مقاييس "جتمان" بشـكل جيد للحصول على معلومات موضوعية أو في الحالات التي يكون فيها الضرورة المنطقية هي أن الاسـتجابة الإيجابية لمسـتوى واحد من التسلسـل الهرمي تعني اسـتيفاء معايير جميع المستويات الأدنى من هذا التسلسل الهرمي. وتصبـح الأمور أكثر غموضـاً عندما لا تكـون الظاهرة محل الاهتمام ملموسـة. وفي حالة مقياسـنا الافتراضي عن الطموح الأبوي، على سـبيل المثال، قد لا يكـون الترتيب موحدًا بين الأفراد. في حين أن تدخين ٢٠ سيجارة في اليوم تعني دائماً تدخين أكثر من ١٠سجائر، والردود عـلى البنديـن رقم (٣) ورقم (٤) في مثال مقياس الطموح الأبوي قد لا تتوافق دائماً مع نمط الترتيب في مقياس «جتمان». فعلى سبيل المثال، قد يوافق الشخص على البند رقم (٣) ولكنه لا يوافق على البند رقم (٤). ومن الطبيعي، أن تنطوي الموافقة على البند رقم (٣) الموافقة على البند رقم (٤)، ولكن إذا رأى المجيب أن النجاحَ عامل معقد أو مركب يعمل في آن واحد كمساعد وعائق أمام السعادة، فإن ذلك قد يسفر عن نمط غير عادي من الاستجابات.

مثـل مقياس «ثورسـتون»، فإن مقياس «جتمان» بلا شـك له مكانتـه، ولكن إمكانية تطبيقهـما تبـدو محدودة إلى حـد ما. ومع كلا النهجـين أو النوعين مـن المقاييس، غالباً مـا تفوق المسـاوئ والمصاعب المزايا. ومن المهم أيضاً التأكيـد من جديد على أن نظريات القياس التي تمت مناقشـتها حتى الآن لا تنطبق دائماً على هذه الأنواع من المقاييس. ومن المؤكد أن افتراض وجود علاقات سـببية قويـة بنفس القدر بين المتغير الكامن وكل بند من البنود لن ينطبق على مقياس «ثورستون» أو مقياس «جتمان». يصف «نونالي وبرنشتاين» Nunnally and Bernstein (1994) بإيجاز بعضاً من النماذج المفاهيمية التي تستند إليها هذه المقاييس. أما بالنسبة للحالات التي تكون فيها البنود المرتبة ملائمة بشكل خاص، فإن النماذج المستندة إلى نظرية استجابة البند (التي تمت مناقشتها في الفصل السابع) قد تكون خياراً مناسباً، على الرغم من أن تنفيذ هذه الأساليب يمكن أن يكون مرهقاً للغاية.

المقاييس ذات البنود متساوية الترجيح أو الوزن:

تتناسب نماذج القياس التي تمت مناقشتها سابقًا بشكل أفضل مع المقاييس التي تتكون من بنود تعادل الى حد ما «أجهزة كشف» للظاهرة محل الاهتمام - أي إنها متكافئة بدرجة أو بأخرى (ولكنها ليسـت متكافئة بالضرورة بالمعنـى الدقيق لنموذج الاختبارات المتوازية «المتكافئـة»). وهي مؤشرات غير كاملة لظاهرة مشـتركة يمكـن دمجها عن طريق الجمع البسيط في مقياس يتمتع بالثبات إلى حد مقبول.

وهنـاك ميزة واحـدة جذابة للمقاييس من هذا النوع هـي أن كل بند من البنود على حدة يمكن أن يكون له مجموعة متنوعة من أشـكال وتنسـيقات خيارات الاستجابة، وهذا يتيح لمطور المقياس قدراً كبيراً من المدى أو الطول في إنشـاء مقياس مناسـب بشكل مثالي للغرض المحدد. وسيتم أدناه بحث بعض القضايا والمشكلات العامة المتعلقة بشكل وصيغة الاستجابة، وكذلك بحث مزايا وخصائص بعض أشكال وصيغ الاستجابة النموذجية.

كم عدد فئات الاستجابة؟

تتكـون معظـم بنود المقياس مـن جزأين هما: جذع أو سـاق، وسلسـلة من خيارات الاسـتجابة (السلم). على سـبيل المثال، قد يكون الساق لكل بند فقرة توضيحية وتفسيرية

مختلفة يعبر عن رأي ما، وقد تكون خيارات الاستجابة المصاحبة لكل ساق عبارة عن سلسلة من الأوصاف التي تشير إلى قوة الاتفاق مع البند. دعونا نركز الآن على خيارات الاستجابة - وعلى وجه التحديد، عدد الخيارات التي ينبغي أن تكون متاحة للمجيب. تسمح بعض أشكال وصيغ استجابة البنود للشخص بوجود عدد لا نهائي أو كبير جدًا من الخيارات، في حين أن بعضها الآخر يحد من الاستجابات المحتملة. تخيل، على سبيل المثال، تدريج الاستجابة لمقياس الغضب الذي يشبه مقياس الحرارة، تمت معايرته من «لا غضب على الإطلاق» في قاعدة المقياس إلى «الغضب التام الذي لا يمكن السيطرة عليه» في قمته. ويمكن تقديم مجموعة من توصيفات الحالة أو الموقف إلى المجيب، مصحوبة كل منها بنسخة من المقياس، ويطلب منه أن يبين، عن طريق التظليل في جزء من المقياس، مدى أو مقدار الغضب الذي تثيره الحالة أو الموقف. تسمح هذه الطريقة بالقياس المتصل «المستمر» للغضب فعلياً. وقد تطلب طريقة بديلة من المجيب أن يبين، باستخدام رقم يتراوح بين درجة واحدة و١٠٠ درجة، مدى ومقدار الغضب الذي يسببه له كل موقف أو حالة. وهذا يوفر العديد من الاستجابات المنفصلة. وبدلاً من ذلك، قد يؤدى الشكل أو التنسيق إلى تقييد خيارات الاستجابة إلى عدد قليل من الخيارات، مثل «لا شيء» أو «قليلاً» أو «مقدار معتدل أو متوسط» و«كثيراً» أو إلى اختيار ثنائي بسيط بين «غاضب» و«غير غاضب».

ما هي المزايا النسبية لهذه البدائل؟ أحد خصائص النوعية أو الجودة المرغوب فيها لتدريج القياس هو الاختلاف والتباين. فالمقياس لا يمكن أن يكون متبايناً إذا لم يختلف أو يتغير. وإذا فشل مقياس ما في التمييز بين الاختلافات في السمة الأساسية، فإن ارتباطاته بالمقاييس الأخرى ستكون مقيدة وسوف تكون فائدته محدودة. ومن بين الطرق لزيادة فرص التباين هي إدراج وتضمين المقياس عدداً كبيراً من البنود. وثمة خيار آخر يتمثل في توفير خيارات عديدة للاستجابة في إطار البنود. إذا اضطرت الظروف الباحث إلى أن يقتصر على سؤالين فقط يتعلقان بقياس الغضب، على سبيل المثال، فقد يكون من الأفضل السماح للمجيبين بمزيد من حرية التعبير في وصف مستوى غضبهم. افترض أن البحث يتعلق بتطبيق سياسات عدم التدخين في بيئة العمل. ولنفترض كذلك أن الباحثين يريدون تحديد العلاقة بين هذه السياسة والغضب. إذا كانت تقتصر على سؤالين فقط (على سبيل المثال، «ما مقدار الغضب الذي تشعر به عندما يتم حظرك من التدخين؟» و«ما مقدار الغضب

الذي تشـعر بــه عندما تتعرض للآخرين ممــن يدخنون في مكان العمــل؟»)، وقد يحصل الباحثــون على معلومات أكثر فائدة من صيغ الاســتجابة التي تســمح للأشــخاص بتدريج اســتجابة أكثر من صيغ الاســتجابة الثنائية (على سبيل المثال، «غاضب» و«ليس غاضب»). فعلى ســبيل المثال، قد يكشف مقياس من «صفر» إلى «١٠٠» عن اختلافات كبيرة في ردود الفعل إزاء هذه الحالات والمواقف كما قد يؤدي إلى تباين جيد في مقياس مكون من بندين. ومــن ناحية أخرى، إذا سُــمح لفريق البحث بإدراج «٥٠» ســؤالاً عــن التدخين والغضب، فإن المؤشرات البســيطة «الغاضبة» مقابل «غير الغاضبة» قد تسفر عن تباين كافٍ عندما تضــاف أو يتم تجميع البنــود للحصول على درجة المقياس. وفي الواقع، فإن مواجهة المزيد من خيارات الاســتجابة بشأن كل ســؤال من الأسئلة الخمسين قد تؤدي إلى إرهاق المجيبين أو تتسبب في إزعاجهم ومللهم، مما يقلل من ثبات إجاباتهم.

وثمة مســألة أخرى تتعلق بعدد خيارات الاستجابة تتمثل في قدرة المجيبين على التمييز بشــكل كبير. ما مدى دقــة التمييز الذي يمكن أن يقوم به الشــخص المثالي؟ ومن الواضح أن هذا يتوقف على ما يتم قياســه. وهناك عدد قليل من الأشياء يمكن تقييمها حقاً في ٥٠ فئة منفصلة، على ســبيل المثال. مع عرض هذه الخيارات العديدة، قد لا يمكن للعديد من المجيبين استخدام سوى تلك الخيارات المقابلة لمضاعفات العدد خمسة أو عشرة، مما يقلل بشكل فعال عدد الخيارات إلى عدد قليل لا يتجاوز خمسة. وقد لا تعكس الاختلافات بين استجابة من ٣٥ إلى ٣٧ الفرق أو الاختلاف الفعلي في الظاهرة التي يتم قياسها. ولا يكسب هذا النوع من الدقة الزائفة إلا القليل. وعلى الرغم من أن تباين المقياس قد يزداد، فإنه قد يكون الجزء العشوائي الذي يتزايد (أي الخطأ) وليس الجزء المنتظم الذي يعزى إلى الظاهرة الأساسية. وهذا، بطبيعة الحال، لا يقدم أو يحقق أي فائدة.

وفي بعض الأحيان، تعتمد قدرة المجيب على التمييز بشكل مفيد بين خيارات الاستجابة عــلى الصياغة المحــددة لتلك الخيــارات أو موضعها (مكانها) المادي. وقــد تؤدي مطالبة المجيــب بالتمييز بين التوصيفات الكمية الغامضة، مثل «عدة» و«قليل» و«كثير»، إلى خلق مشــاكل. في بعض الأحيان، يمكن تقليل الغموض عن طريق ترتيب خيارات الاستجابة على الصفحــة. ويبــدو أن المجيبين غالبًا مــا يفهمون ما هو مرغوب فيــه عندما تُعرض عليهم سلسلة متصلة واضحة. وبالتالي، فإن الطلب مثل:

كثيراً	بعض الأحيان	قليلاً	قليلاً جداً	أبداً

وقد يعني هذا ضمنياً أن الخيار «بعض الأحيان» هو أكثر من خيار الاستجابة «قليلاً» بسبب ترتيب هذه البنود. ومع ذلك، إذا كان من الممكن إيجاد صفة لا لبس فيها تمنع المجيبين من وضع افتراضات تستند إلى الموضع أو المكان على طول سلسلة متصلة، فإن هذا يكون أفضل بكثير. وفي بعض الأحيان، قد يكون من الأفضل أن يكون لديك عدد من خيارات الاستجابة أقل من تلك التي تكون غامضة. لذلك، على سبيل المثال، قد يكون من الأفضل في المثال السابق حذف إما «بعض الأحيان» أو «قليلاً» ويكون لديك أربعة خيارات بدلاً من خمسة. وأسوأ الظروف هي الجمع بين الكلمات الغامضة ومواقع الصفحة الغامضة. فكر في المثال التالي:

مفيد جداً Very Helpful	غير مفيد جداً Not Very Helpful
مفيد إلى حدٍ ما Somewhat Helpful	غير مفيد على الأطلاق Not at All Helpful

ومن الصعب جداً التمييز بين مصطلحات مثل إلى حدٍ ما somewhat "وليس إلى حدٍ كبير not very" في ظل أفضل الظروف. ومع ذلك، فإن ترتيب خيارات الاستجابة هذه كما تظهر أعلاه يجعل الأمور أكثر سوءاً. إذا قام المجيب بقراءة العمود الأول ثم أسفل العمود الثاني، فيبدو أن تعبير "إلى حدٍ ما somewhat" يمثل قيمة أعلى من تعبير "ليس إلى حدٍ كبير not very". ولكن إذا قرأ المجيب الصف الأول ثم عبر الصف الثاني، فإن الترتيب الضمني لهذين الاستجابتين الواصفتين على طول سلسلة متصلة يتم عكسه. ونظرًا للغموض الذي يكتنف كلاً من الترتيب اللغوي والترتيب المكاني، يمكن للأفراد أن يعطوا معاني مختلفة للخيارين اللذين يمثلان قيمتين معتدلتين، وكنتيجة لذلك، فإن الثبات سيتأثر.

وثمة مسألة أخرى تتمثل في قدرة الباحث واستعداده لتسجيل عدد كبير من القيم لكل بند. إذا تم استخدام طريقة مقياس الحرارة التي تم وصفها سابقاً لتحديد الاستجابات الكمية على مقياس الغضب، فهل سيحاول الباحث بالفعل تسجيل درجات دقيقة لكل استجابة؟ ما هي الدقة المناسبة؟ هل يمكن قياس المنطقة المظللة في حدود ربع بوصة؟ سنتيمتر؟ مليمتر؟ إذا تم استخراج بعض البيانات الخام فقط - مثل الثلث الأدنى أو الأوسط أو الأعلى - من المقياس، فما هي النقطة التي كانت عندها المطالبة بمثل هذه الاستجابة الدقيقة؟

وهناك مســألة أو مشــكلة أخرى على الأقل تتعلق بعدد الردود أو الاســتجابات. على افــتراض الســماح بعدد قليل من الإجابــات المنفصلة لكل بند، هل ينبغــي أن يكون عدد الردود أو الاستجابات فردياً أم زوجياً؟ ومرةً أخرى، يعتمد هذا على نوع السؤال، ونوع خيار الاســتجابة، وغرض أو هدف الباحث. وإذا كانت خيارات الاســتجابة ثنائية القطب، حيث يشــير أحد القطبين إلى عكس القطب الآخر (على ســبيل المثال، اتجــاه إيجابي قوي مقابل اتجاه ســلبي قوي)، فإن عددًا فرديًا من الردود والاســتجابات يسمح (بالمراوغة) بالغموض واللبس (على ســبيل المثال، "لا أتفق ولا أعارض neither agree nor disagree") أو عدم اليقــين أو عدم التأكد (على ســبيل المثــال، "غير متأكد")؛ وعادةً لا يســمح العدد الزوجي للاســتجابات بذلك. وينطوي العدد الفردي للاستجابات على نقطة "محايدة" مركزية (على ســبيل المثال، لا تقييم إيجابي ولا ســلبي). من ناحية أخرى، هناك عدد زوجي من الردود أو الاســتجابات يجبر المجيب على أن يقدم على الأقــل التزاماً ضعيفاً في اتجاه أحد الطرفين أو الطــرف الآخر(مثــل الاختيار القسري بين التقييم الإيجابي إلى حد ما أو التقييم الســلبي إلى حــد ما باعتبــاره الرد الأقل تطرفاً). ولا يكون أي من الشــكلين متفوقاً بالضرورة. وقد يرغــب الباحــث في منع (المراوغة) الغموض واللبس إذا كان هناك شــعور بأن الأشــخاص سيختارون اســتجابة محايدة كوســيلة لتجنب الاختيار. ففي الدراسات المتعلقة بخيارات المقارنة الاجتماعية، على ســبيل المثال، قد يرغب الباحثون في إجبار الأشــخاص على التعبير عن تفضيلهم للحصول على معلومات حول الشــخص الــذي يتمتع بقدر أكبر من المزايا أو بقــدر أقل من المزايا. لننظر في هذين الشــكلين البديلين، الأول الذي تم اختياره لدراســة المقارنات الاجتماعية بين المصابين بالتهاب المفاصل (DeVellis et al., 1990):

١- هل تفضل معلومات حول:

(أ) المرضى الذين يعانون من التهاب المفاصل أسوأ مما لديك.

(ب) المرضى الذين يعانون من التهاب المفاصل أكثر اعتدالاً مما لديك.

٢- هل تفضل معلومات حول:

(أ) المرضى الذين يعانون من التهاب المفاصل أسوأ مما لديك.

(ب) المرضى الذين يعانون من التهاب المفاصل بنفس القدر الذي تعاني منه.

(ج) المرضى الذين يعانون من التهاب المفاصل أكثر اعتدالاً مما لديك.

قد يسمح خيار محايد مثل رقم ٢ (ب) بالمراوغة (اللبس والغموض) غير المرغوب فيها. وقد يكون من المطلوب والمستحسن أيضاً وجود نقطة محايدة. في دراسة تقييم أي نوع من المخاطر (على سبيل المثال، الملل مقابل الخطر) الذي يفضله الناس، قد تكون نقطة الوسط ضرورية أو حاسمة. قد يغير الباحث فرصة أو شدة الضرر عبر عدة خيارات بين نشاط آمن وممل ونشاط مثير ومحفوف بالمخاطر. ويمكن عندئذ استخدام النقطة التي يكون فيها المجيب ملتبساً إلى أبعد حد بشأن المخاطرة بالنشاط الأكثر إثارة كمؤشر للمخاطرة:

يرجى الإشارة إلى تفضيلك النسبي للنشاط (أ) أو النشاط (ب) من بين البدائل المدرجة أدناه من خلال وضع دائرة حول العبارة المناسبة بعد وصف النشاط (ب).

النشاط (أ): قراءة كتاب إحصائي (لا توجد فرصة لحدوث إصابة شديدة).

١- النشاط(ب): القيام برحلة جوية في طائرة ركاب صغيرة (احتمال ضئيل جداً بحدوث إصابة بالغة).

أفضل بشدة النشاط (أ)	أفضل بشكل معتدل النشاط (أ)	لا أفضل أي منهما	أفضل بشكل معتدل النشاط (ب)	أفضل بشدة النشاط (ب)

٢- النشاط (ب): القيام برحلة على متن طائرة صغيرة في قمرة القيادة (فرصة طفيفة للإصابة الشديدة).

أفضل بشدة النشاط (أ)	أفضل بشكل معتدل النشاط (أ)	لا أفضل أي منهما	أفضل بشكل معتدل النشاط (ب)	أفضل بشدة النشاط (ب)

٣- النشاط (ب): القفز بالمظلة من طائرة مع مظلة احتياطية (فرصة بشكل معتدل لإصابة شديدة).

أفضل بشدة النشاط (أ)	أفضل بشكل معتدل النشاط (أ)	لا أفضل أي منهما	أفضل بشكل معتدل النشاط (ب)	أفضل بشدة النشاط (ب)

٤- النشـــاط (ب): القفز بالمظلة من طائرة بدون مظلــة احتياطية (خطر كبير من الإصابة الشديدة).

أفضل بشدة النشاط (أ)	أفضل بشكل معتدل النشاط (أ)	لا أفضل أي منهما	أفضل بشكل معتدل النشاط (ب)	أفضل بشدة النشاط (ب)

٥- النشـــاط (ب): القفز من طائرة بـــدون مظلة ومحاولة الهبوط على هدف ناعم (إصابة خطيرة مؤكدة تقريباً).

أفضل بشدة النشاط (أ)	أفضل بشكل معتدل النشاط (أ)	لا أفضل أي منهما	أفضل بشكل معتدل النشاط (ب)	أفضل بشدة النشاط (ب)

وبـــصرف النظر عن المزايـــا أو الالتزامات الأخرى لهذا النهج أو الأســـلوب، فإنه يتطلب بوضوح أن تتضمن خيارات الاستجابة نقطة وسط أو نقطة منتصف.

أنواع محددة (معينة) من أشكال وتنسيقات الاستجابة:

تظهر بنود المقياس في مجموعة متنوعة من الأشـــكال والصيغ المثيرة. ومع ذلك، هناك عدة طرق لعرض وتقديم البنود التي يتم اســـتخدامها على نطاق واســـع وأثبتت نجاحها في تطبيقات متنوعة. ونناقش أدناه بعض هذه المسائل.

مقياس ليكرت:

يٌعد مقياس «ليكرت» Likert scale أحد الأشـــكال والصيغ الأكثر شـــيوعاً للبنود. وعند اســـتخدام مقياس "ليكرت"، يتم عرض البند كجملة تفســـيرية، تليها خيارات استجابة تشير إلى درجات مختلفة من الاتفاق مع العبارة أو تأييدها. (في الواقع، كان المثال الســـابق عن المخاطرة يســـتخدم شـــكل وصيغة اســـتجابة مقياس ليكرت). وتبعـــاً للظاهرة التي يجري دراستها وأهداف الباحث، فقد يصاحب كل بند إما عدد فردي أو عدد زوجي من خيارات الاســـتجابة. وينبغي صياغة خيارات الاستجابة بحيث تكون لها فترات متساوية تقريباً فيما يتعلـــق بالاتفاق. وهذا يعني أن الفرق في الاتفاق بين أي زوج متجاور من الإجابات يجب

أن يكـون مماثـلاً تقريباً لأي زوج متجاور آخر من خيارات الاسـتجابة. ومن الممارسـات الشـائعة إدراج ستة ردود أو اسـتجابات محتملة هي: «غير موافق بشدة»، و«غير موافق إلى حد ما»، و«غير موافق بشكل معتدل أو متوسط»، و«موافق بشكل معتدل أو متوسط»، و«موافق إلى حد ما»، و«موافق بشـدة». وهذا يشـكل سلسـلة متصلة من عدم الاتفاق بشـدة إلى الاتفاق بشـدة. يمكن أيضاً إضافة نقطة وسـط أو منتصف محايدة. وتشـمل الخيارات الشـائعة لنقطة الوسـط «لا أوافق ولا أعارض» و«موافق وغير موافق بالتساوي (سـيان)». فهناك مجال مشروع للمناقشة فيما يتعلق بتكافؤ هاتين النقطتين المتوسطتين. الأول يعنـي ضمنياً عدم الاكتراث واللامبالاة، في حين يشـير الأخـير إلى وجود جاذبية قوية ولكنها متسـاوية لكل من الاتفاق وعدم الاتفاق. وقد يكون الأمر جيدًا أن معظم المجيبين لا يركزون كثيراً على الفروق الدقيقة في اللغة، بل يعتبرون مجرد خيار الاستجابة المعقول في وسط النطاق كنقطة وسط أو منتصف، بصرف النظر عن صياغته الدقيقة.

يسـتخدم مقيـاس «ليكـرت» على نطـاق واسـع في أدوات قيـاس الآراء والمعتقدات والاتجاهـات. ومن المفيد غالباً أن تكون هذه العبارات قوية إلى حدٍ ما (وإن لم تكن قوية للغاية) عند اسـتخدامها في شكل وصيغة مقياس ليكرت. ومن المفترض أن يتم التعبير عن الاعتدال في الرأي في اختيار خيار الاسـتجابة. على سبيل المثال، العبارات التالية: «يتجاهل الأطبـاء عمومًا ما يقولـه المرضى»، و«في بعض الأحيان، لا يـولي الأطباء القدر من الاهتمام الـذي ينبغي لتعليقات المرضى»، و«مـرة واحدة من حين لآخر، قد ينسى الأطباء أو يفوتهم شيئ ما أخبرهم به المريض» هي تعبر عن آراء قوية وآراء معتدلة وآراء ضعيفة، على التوالي، فيـما يتعلق بعدم اهتمام الأطباء بتعليقـات المرضى. أيهم يكون أفضل لمقياس «ليكرت»؟ في نهايـة المطاف، بطبيعـة الحال، فإن الذي يعكس الاختلافات الحقيقية في الرأي بشـكل أكثر دقة يكون هو الأفضل. عند اختيار مدى قوة الصياغة في تجمع أوَّلي للبنود، قد يكون من المفيد أن يتسـاءل الباحث عن «الطريقة التي من المحتمل أن يجيب بها الناس الذين لديهـم مقادير أو نقاط قوة مختلفة في السـمة المعنية» وفي حالـة الأمثلة الثلاثة المقدمة للتو، فإن الباحث قد يسـتنتج أن السـؤال الأخير ربما قد يؤدي إلى التوصل إلى اتفاق قوي على الأرجح من جانب الأشـخاص الذين تندرج آراؤهم على امتداد جزء كبير من السلسـة المتصلة من الإيجابية إلى السـلبية. وإذا ثبت أن هذا الاسـتنتاج صحيح، فلن تؤدي العبارة

الثالثة دورًا جيدًا في التمييز بين الأشــخاص ذوي الآراء الســلبية القوية مقابل الآراء السلبية المعتدلة (المتوسطة).

وبصفــة عامة، قد تؤدي العبارات المعتدلة للغاية إلى التوصل إلى اتفاق أكثر مما ينبغي عند اســتخدامها في مقاييس «ليكرت». وســوف يتفق كثير من الناس بشدة مع عبارة مثل «سلامة وأمن المواطنين أمر مهم». ويمكن للشخص أن يوافق بشدة على مثل هذه العبارة (أي اختيار خيار الاستجابة القصوى) دون إبداء رأي مبالغ فيه. وبطبيعة الحال، فإن العكس هو الصحيح بنفس القدر. وقد يجد الأشــخاص الذين لديهم أي آراء أكثر تطرفاً أنفسهم في خلاف (عدم اتفاق) مع عبارة قوية للغاية (على ســبيل المثال، «إن ملاحقة مرتكبي الأفعال غير المشروعة ومعاقبتهم أهم من حماية حقوق الأفراد»). ومن بين العبارتين (الأكثر اعتدالاً والأكثر تطرفاً)، فإن الأولى قد تكون هي المشكلة الأكبر لسببين. أولاً، غالباً ما نميل إلى كتابة عبــارات لا تسيء إلى الناس. إن تجنب الهجوم قــد يكون فكرة جيدة. ومع ذلك، فإنه قد يقودنا إلى تفضيل البنود التي سيجدها الجميع تقريباً مقبولة. وهناك سبب آخر يدعو إلى الحذر من البنود المعتدلة جداً، وهو أنها قد تمثل غياب أو عدم وجود الاعتقاد أو الرأي. أما البند الثالث من بنودنا الخاصة بالطبيب الذي لا ينتبه لتعليقات المرضى في الفقرة السابقة فلم يبين وجود اتجاه إيجابي بقدر ما يدل على عدم وجود اتجاه غير إيجابي. وقد لا تتناسب بنود من هذا النوع بدرجة كافية مع هدف البحث لأننا غالباً ما نهتم بوجود ظاهرة ما أكثر مما نهتم بعدم وجودها.

تتمثــل إحدى الطرق المفيدة لمعايرة مدى قــوة أو اعتدال صياغة عبارة ما في القيام بما يلي: تخيل المســتجيب العادي الذي يكون هدفــاً لتطبيق المقياس. حاول أن تتخيل كيف سيســتجيب هذا الشــخص على بنود مختلفة من حيث القوة. والآن، فكر في نوع صياغة البند الذي من المرجح أن يســتخلص إجابة من المجيب العادي الذي كان في مركز خيارات الاســتجابة لمقياس «ليكرت» والذي تخطط لاســتخدامه أو بالقرب منه. لذلك، على سبيل المثــال، إذا اخترت مقياسًــا ذا تدريج مكون من ٦ نقاط وكان فيه خيار الاســتجابة المركزي (الوسط) هو «غير موافق قليلاً وموافق قليلاً»، فإنك تريد تكوين بند يستخلص إحدى هذه الاســتجابات من مجيب عادي في المجتمع الإحصائي محل الاهتمام. وينبغي أن يكون مثل هذا البند قادرًا على اســتيعاب الأشــخاص الذين تكون وجهــات نظرهم (آراؤهم) إما أقل

أو أكثر قوة من المجيب العادي أو المتوسـط، وبالتالي ينبغي من الناحية النظرية أن يكون لديهم تباين كبير في الدرجة الملاحظة والدرجة الحقيقية. وبفضل التباين بين المجيبين، ستتاح للبند فرصة أفضل للارتباط بشـكل جيد مع البنود الأخرى (لأن التباين المشـترك «التغاير» مرتبط بمدى التباين) وبالتالي سيكون لديه القدرة على تعزيز ثبات المقياس. وعلى النقيض من ذلك، فإن البند الذي من المحتمل أن تنتج عنه اسـتجابات شـديدة «متطرفة أو مبالغ فيها»، مثل عدم الموافقة بشـدة أو الموافقة بشـدة، من المجيب العادي، سوف يؤدي عملاً ضعيفاً في التمييز بين مختلف الفئات الكاملة من المجيبين.

باختصـار، يجـب أن يوضح بند مقياس "ليكرت" الجيد الـرأي أو الاتجاه أو المعتقد أو أي مفهوم بنائي آخر قيد الدراسـة بعبارات واضحة. وليس من الضروري ولا من المناسـب أن يمتـد هذا النوع من المقياس إلى نطاق التأكيـدات الضعيفة إلى تأكيدات قوية للمفهوم البنائي. توفر خيارات الاستجابة الفرصة للتدرجات.

فيما يلي أمثلة للبنود بصيغ وأشكال استجابة مقياس «ليكرت»:

١- تعتبر ممارسة التمارين الرياضية عنصراً أساسياً في نمط حياة صحي.

١	٢	٣	٤	٥	٦
غير موافق بشدة	غير موافق إلى حد ما	غير موافق بشكل معتدل	موافق بشكل معتدل	موافق إلى حد ما	موافق بشدة

٢- ينبغي أن تكون مكافحة تعاطي المخدرات أولوية وطنية عليا أو قصوى.

١	٢	٣	٤	٥
صحيح تماماً	صحيح في الغالب	صحيح وغير صحيح على حد سواء	غير صحيح في الغالب	غير صحيح تماماً

القياس التفاضلي الدلالي (تباين الدلالات اللفظية):

ترتبــط طريقة القيــاس التفاضلي الدلالي بشــكل رئيسي ببحوث الاتجــاه التي أجراها «أوســقود وزملاؤه» (على ســبيل المثــال. Osgood & Tannenbaum, 1955). وعادةً، ما يســتخدم تباين الدلالات اللفظية في الإشــارة إلى واحد أو أكثر من الدوافع أو المحفزات. ففي حالة الاتجاهات، على سبيل المثال، قد يكون الحافز أو الدافع هو مجموعة من الناس، مثل مندوبي مبيعات الســيارات. ويتبع تحديد الدوافع والحوافز المســتهدفة قائمة بأزواج الصفــات. يمثل كل زوج طرفي نقيض من سلســلة متصلة، محددة بالصفات (على ســبيل المثال، أمين أو صادق وغير أمين). كما هو موضح في المثال أدناه، هناك عدة نقاط (سطور) بين الصفات التي تشكل خيارات الاستجابة:

مندوب مبيعات السيارات:								
أمين (صادق)	-	-	-	-	-	-	-	غير أمين
هادئ (لطيف)	-	-	-	-	-	-	-	مزعج

في الأساس، تمثل الخطوط الفردية (سبعة وتسعة أرقامٍ مشتركة) نقاطًا على طول السلسلة المتصلة المحددة بواسطة الصفات. يضع المجيب علامة أو إشارة على أحد الخطوط للإشارة إلى النقطة الموجودة على طول السلســلة المتصلة التي يتســم بها تقييمه للحافز أو الدافع. على ســبيل المثال، إذا كان شخص ما يعتبر مندوب مبيعات السيارات غير أمين للغاية، فقد يختــار هو أو هي الخط الأقــرب إلى تلك الصفة. يمكن التعبير عــن وجهات النظر والآراء المتطرفــة أو المعتدلــة من خلال اختيــار الخط الذي يجب وضع علامــة علية. بعد تقييم الدوافــع "الحوافز" فيما يتعلق بزوج الصفة الأول، ينتقل الشــخص إلى أزواج صفة إضافية مفصولة بخطوط.

إن الصفــات التي يختارها الشــخص قد تكون إمــا ثنائية القطــب أو أحادية القطب، اعتــمادًا، كــما هو الحال دائمًا، على المنطق الذي تســتند إليه الأســئلة البحثية التي يهدف المقيــاس إلى معالجتهــا. كل الصفــات ثنائيــة القطب تعــبر عن وجود ســمات متناقضة،

مثل الودية friendly والعدائية hostile. وتشـير أزواج الصفات أحادية القطب إلى وجود سمة واحدة وعدم وجودها (غيابها)، مثل الودية friendly وغير الودية.

كما هو الحال مع مقياس "ليكرت"، يمكن أن يكون شـكل وصيغة الاسـتجابة التفاضلية الدلالية متوافقًا إلى حدٍ كبير مع النماذج النظرية التي تم عرضها في الفصول السابقة من هذا الكتاب. يمكن كتابة مجموعات من البنود للاسـتفادة من نفس المتغير الأسـاسي. على سبيل المثال، يمكن إضافة البنود التي تستخدم نقاط النهاية التالية، جدير بالثقة / غير جدير بالثقة، وعـادل / وغير عـادل، وصادق / غير صادق كنقاط نهاية إلى العبارة الأولى في المثال السـابق بحيث تشـكل مقياس "الشرف أو الأمانة". يمكن تصور مثل هذا المقياس على أنه مجموعة من البنود التي تشترك في متغير كامن مشترك هو (الشرف أو الأمانة) وتتوافق مع الافتراضات التي تمت مناقشتها في الفصل الثاني. وبناءً على ذلك، يمكن إضافة وتحليل العشرات من البنود كل على حدة «لمقياس الأمانة» على النحو المبين في قسم لاحق فيما يتعلق بتقييم البنود.

التناظر البصري:

هناك تنسـيق آخر لشكل وصيغة البند يشـبه إلى حدٍ ما القياس التفاضلي الدلالي وهو مقياس التناظر البصري. يقدم شـكل وصيغة الاسـتجابة هذا للمستجيب خطًا متصلاً بين زوج مـن التوصيفـات الذي يمثل نهايات على طرفي نقيض (متعاكسـة) لسلسـلة متصلة. ويُطلب من الشخص الذي يكمل البند وضع علامة عند نقطة على الخط تمثل رأيه أو خبرته أو معتقده أو أي شيء يتم قياسه. إن مقياس التناظر البصري، كما يوحي مصطلح التناظرية في الاسـم، هو مقياس مسـتمر أو متصل. يقوم فيه الباحث بتحديد دقة التمايز في تعيين الدرجات إلى النقاط على المقياس. وقد تم في وقت سابق مناقشة بعض مزايا وعيوب شكل أو صيغة الاستجابة المستمرة. وهناك مسألة أو مشكلة إضافية لم تتم إثارتها في ذلك الوقت تتعلق بالاختلافات المحتملة في تفسير المساحة المادية من حيث صلتها بالقيم على المتصل. قد لا تعني العلامة الموضوعة في نقطة معينة على طول الخط نفس الشيء بالنسبة لأشخاص مختلفين، حتى عندما يتم تمييز نقاط النهاية للخط بشكل مماثل لجميع المجيبين. فكر في مقياس تناظر بصري للألم مثل التالي:

لا ألم على الإطلاق .... أسوأ ألم واجهته أو جربته.

هل تشـير الاسـتجابة في منتصف المقياس إلى الألم الذي يصيب نصـف الوقت، أو الألم المستمر بنصف الشدة المحتملة، أو شيء آخر بالكامل؟ جزء من المشكلة مع قياس الألم هو أنه يمكن تقييمه على أبعاد متعددة، بما في ذلك التكرار، والشـدة، والمدة. كما أن ذكريات أسـوأ ألم قد تعرض له شـخص ما من المرجح أن تكون مشـوهة. والواقع أن المقارنات بين الأفراد تزداد تعقيداً بفعل حقيقة مفادها أن الأشخاص المختلفين ربما قد تعرضوا لمستويات مختلفة من «أسـوأ ألم». وبطبيعة الحال، لا شك أن بعض هذه المشاكل تكمن في الظاهرة المستخدمة في هذا المثال - وهي الألم (انظر «كيف» Keefe, 2000، للحصول على مناقشة ممتـازة لقيـاس الألم) - وليس مع المقياس في حد ذاته. ومع ذلك، يمكن أن توجد مشـكلة تعيـين للقيم بطريقة غير اعتيادية (عشـوائية) على طول مقيـاس تناظري بصري لظواهر أخرى أيضاً.

ومن المميزات الرئيسـية لمقاييس التناظر البصري أنها قد تكون حساسـة للغاية (ماير، Mayer, 1978)، وهذا من شـأنه أن يجعلها مفيدة بشـكل خـاص في قياس الظواهر قبل وبعد وقوع بعض الأحداث المتداخلة، مثل التدخل أو المعالجة التجريبية، التي تمارس تأثيراً ضعيفاً نسبياً. على سبيل المثال، قد لا يؤدي التوبيخ الخفيف أو المتوسط في سياق المعالجة التجريبيـة إلى إحـداث تحول في مقياس لاحترام الذات ذي تدريـج خماسي النقاط. ولكن ربمـا قد يحدث تحول دقيق ولكنه منهجي أو منتظم يؤدي إلى انخفاض القيم على مقياس تناظري بصري بين الناس في حالة «التوبيخ» لهذه التجربة الافتراضية. وقد تكون الحساسية أكثر فائدة عند فحص ودراسـة التغيرات على مر الزمن داخل نفس الفرد وليس بين الأفراد (مايـر، Mayer, 1978). وقد يكون الأمـر كذلك لأنه في الحالة الأولى، لا يوجد خطأ إضافي بسبب الاختلافات الخارجية بين الأفراد.

وهناك ميزة أخرى محتملة لمقاييس التناظر البصري عند تكرارها بمرور الوقت، وهي أنه من الصعب أو المستحيل على الأشخاص ترميز «تشفير» إجاباتهم السابقة بدقة. بالاستمرار في المثـال الوارد في الفقرة السـابقة، ربما لا يجد شـخص ما صعوبة كبـيرة في تذكر أي من الخيارات الخمسـة المرقمة بالنسـبة لبند تقدير الذات الذي كان قد اختاره أو اختارته من قبل استجابة لشكل أو صيغة متعددة الاستجابات مثل مقياس «ليكرت». إلا أنه ما لم يتم اختيار إحدى نقاط النهاية لمقياس تناظر بصري، سيكون من الصعب أن نتذكر بدقة المكان

الذي تم فيه وضع علامة على طول خط غير مميز. وقد يكون ذلك مفيدًا إذا كان الباحث يشــعر بالقلق مــن أن المجيبين قد ينحازون إلى الظهور بمظهر متســق مــع مرور الوقت. ومن المفترض أن يختار الأشــخاص الذين لديهم دوافع ليكونوا متســقين نفس الاســتجابة بعد التعرض لتدخل تجريبي كما كان الحال قبل التعرض لمثل هذا التدخل. غير أن شــكل أو صيغة التناظر البصري يســتبعد هــذا الاحتمال. إذا كانت اســتجابات ما بعد المعالجة التجريبية قد خرجت بشــكل ثابت (أي في نفس الاتجاه عادةً) عن الاستجابة السابقة قبل المعالجة بالنســبة للموضوعات التجريبية وبشــكل عشــوائي لعناصر الضبط والتحكم، فإن اختيار مقياس تناظر بصري قد يكون ســاهم في الكشــف عن ظاهرة دقيقة كانت أساليب أخرى تخفيها.

غالبًا ما تســتخدم مقاييس التناظر البصري كمقاييس أحادية البند. وهذا له عيب كبير يتمثل في اســتبعاد أي تحديد للاتســاق الداخلي. ومع المقياس أحادي البند، فإنه لا يمكن تحديد الثبات إلا من خلال استخدام طريقة إعادة الاختبار test-retest الموضحة في الفصل الثالث أو بالمقارنة مع مقاييس أخرى لنفس الســمة التي لها خصائص سيكومترية راسخة. وتعاني الطريقة الأولى من مشــكلات تقييمات إعادة الاختبار التي تمت مناقشتها سابقًا، ولا ســيما اســتحالة التمييز بين عدم الاستقرار في عملية القياس وعدم اســتقرار الظاهرة التي يتم قياســها. أما الطريقة الأخيرة فهي في الواقع تقابل (تشــبه) صدق المفهوم البنائي. ومع ذلك، نظرًا لأن الثبات شرط ضروري للصدق، يمكن للمرء أن يســتنتج الثبات إذا كان هناك دليل على الصدق. ومع ذلك، قد تكون الإستراتيجية الأفضل هي تطوير عدة بنود تناظرية بصرية بحيث يمكن تحديد الاتساق الداخلي.

أشكال الاستجابة العددية والعمليات العصبية الأساسية:

تشــير دراســة أجرتها «زورزي وبرفيتليس وأوميليتــا» Zorzi, Priftis, and Umilitá (2002) نُــشرت في مجلة الطبيعة Nature إلى أن بعض خيارات الاســتجابة قد تتوافق مع كيفيــة معالجة الدمــاغ للمعلومات الرقميــة أو العددية. وفقاً لما ذكــره هؤلاء المؤلفون، فــإن الأرقام المرتبة في تسلســل، كما هو الحال مع مقياس "ليكــرت" النموذجي، تعبر عن الكميــة ليس فقط في قيمها الرقمية أو العددية ولكن في مواقعها. ويقترح هؤلاء المؤلفون

أن الخط البصري للأرقام ليس مجرد تمثيل مناسب بل إنه يتوافق مع العمليات العصبية الأساسية. ولاحظوا أن الأشخاص الذين يعانون من إصابات مختلفة في الدماغ والتي تؤثر على الإدراك المكاني في المجال البصري يقومون بارتكاب أخطاء منهجية في مسائل الرياضيات البسيطة التي يتم تقديمها بصرياً (أو بصورة مرئية). ويرتبط الشذوذ "الانحراف" المكاني ونوع الأخطاء ارتباطاً وثيقاً. فالأفراد الذين لم يتمكنوا من إدراك المجال البصري أو المرئي الأيسر، عندما طُلب منهم الإشارة إلى نقطة الوسط أو المنتصف بين قيمتين معروضتين في ترتيب أو مصفوفة خطية، كانوا يخطئون دوماً "إلى اليمين". على سبيل المثال، عندما سئل الأفراد عن النقطة المتوسطة بين النقطتين "٣" و"٩"، تحولت الأخطاء إلى اليمين (أي إلى القيم الأعلى). وبعكس المقياس من القيم الأعلى إلى القيم الأدنى استمر في إحداث تحولات إلى اليمين (الآن، القيم الأدنى). وعندما تم تقديم نفس المهام في شكل غير مرئي (على سبيل المثال، من خلال طرح السؤال عن متوسط ٣ و٩)، لم يظهر النمط. والواقع، لم يظهر هؤلاء الأفراد أي عجز في أداء الحسابات عندما لم يتم عرضها بصرياً. لم يظهر الأشخاص الخاضعين للضبط دون وجود الشذوذ أو الانحراف البصري نمط التحول لدى الذين يعانون من إصابات أو آفات الدماغ. ويخلص المؤلفون إلى أن عملهم يشكل «دليلاً قوياً على أن خط الأرقام الذهنية هو أكثر من مجرد تعبير مجازي أو تشبيه وأن «التفكير في الأرقام من الناحية المكانية (كما ذكر علماء الرياضيات العظماء) قد يكون أكثر كفاءة لأنه يقوم على أساس التمثيل العصبي الفعلي للأرقام» (ص ١٣٨). وعلى الرغم من أن هذه الدراسة، في حد ذاتها، قد لا تبرر استنتاجات قاطعة وسريعة، إلا أنها توفر أدلة أولية جديرة بالاهتمام على أن تقييم سلسلة خطية من الأرقام قد تتوافق مع الآليات العصبية الأساسية المتضمنة في التقييم الكمي. وإذا كان هذا هو الحال بالفعل، فإن خيارات الاستجابة المعروضة كصف من الأرقام قد تكون لها ميزة خاصة.

الخيارات الثنائية:

يوفر شكل وصيغة استجابة شائع آخر للأشخاص فرصة الاختيار بين الخيارات الثنائية لكل بند. وكانت الأمثلة السابقة لمقياس «ثرستون» Thurstone ومقياس "جتمان" Guttman تستخدم الخيارات الثنائية ("موافق" و"غير موافق")، على الرغم من أن المقاييس التي

تحتوي على البنود متساوية الوزن قد تشتمل أيضاً على خيارات استجابة ثنائية. على سبيل المثال، قد يُطلب من الأشخاص التأشير على جميع الصفات الموجودة في قائمة يعتقدون أنها تنطبق على أنفسهم، أو قد يُطلب منهم الإجابة بـ "نعم" أو "لا" على قائمة ردود الفعل العاطفية التي قد تكون لديهم في بعض المواقف المحددة. وفي كلتا الحالتين، يمكن جمع أو دمج الاستجابات التي تعكس بنود تشترك في متغير كامن مشترك (على سبيل المثال، الصفات مثل "حزين" و"غير سعيد" و"الأزرق" والتي تمثل الاكتئاب) في درجة واحدة لهذا المفهوم البنائي.

ويتمثل أحد أوجه القصور الرئيسية في الاستجابات الثنائية في أن كل بند لا يمكن أن يكون له سوى أقل قدر من التباين (القابلية للتغير). وبالمثل، فإن أي زوج من البنود يمكن أن يكون له مستوى واحد فقط من مستويين من التغاير (التباين المشترك): الاتفاق أو عدم الاتفاق. تذكر من الفصل الثالث أن تباين المقياس الذي يتكون من عدة بنود متساوية الوزن أو الترجيح يساوي تماماً مجموع كافة العناصر في مصفوفة التغاير لكل بند على حدة. وفي حالة البنود ثنائية الاستجابة، فإن كل بند يساهم بقدر ضئيل في هذا المجموع بسبب القيود في التباينات والتغايرات (التباينات المشتركة) المحتملة. والنتيجة العملية لذلك هي وجود حاجة إلى المزيد من البنود للحصول على نفس درجة تباين المقياس إذا كانت البنود ثنائية الاستجابة. ومع ذلك، عادةً ما تكون الإجابة على البنود ثنائية الاستجابة سهلة للغاية. ولذلك، فإن العبء الملقى على عاتق الشخص يكون منخفضاً بالنسبة لأي بند من البنود. على سبيل المثال، يمكن لمعظم الناس أن يقرروا بسرعة ما إذا كانت بعض الصفات هي أوصاف مناسبة لأنفسهم. ونتيجة لذلك، غالباً ما يكون الأشخاص على استعداد لإكمال المزيد من البنود ثنائية الاستجابة أكثر من تلك البنود التي تستخدم شكل أو صيغة تتطلب التركيز على الفروق الدقيقة. وبالتالي، فإن الشكل ثنائي الاستجابة قد يسمح للباحث بتحقيق تباين كافٍ في درجات المقياس عن طريق تجميع المعلومات على المزيد من البنود.

الأطر الزمنية للبند:

ثمة مشكلة أخرى تتعلق بشكل وتنسيق البنود تتمثل في الإطار الزمني المحدد أو الضمني. يناقش كيلي ومكجراث (1988) Kelly and McGrath، في مجلد آخر من هذه السلسلة،

أهميــة النظر والأخــذ في الاعتبار الســمات الزمنية للمقاييس المختلفة. ولن تشــير بعض المقاييــس إلى إطــار زمني، مما يعني وجــود منظور زمني عالمي. على ســبيل المثال، كثيراً ما تتضمن مقاييس موضع التحكم والســيطرة، على بنود تنطوي على إيمان دائم بالعلاقات السببية. وتفترض بعض البنود مثل «إذا اتخذت الإجراءات الصحيحة، فبوسعي أن أبقى في صحة جيدة» (والســتون وآخرون Wallston et al., 1978) أن هذا الاعتقاد مستقر نسبيًا. ويتوافق هذا مع التوصيف النظري لموضع التحكم والســيطرة باعتباره متوسطاً عامًا وليس محددًا لضبط النتائج (على الرغم من حدوث تحول نحو مزيد من الخصوصية والتحديد في مقاييس لاحقة لموضع معتقدات التحكم - مثل ديفاليس وآخرون، DeVellis et al., 1985). وهنــاك مقاييس أخــرى لتقييم الظواهر المؤقتة (العابرة) نســبياً. فالاكتئاب، على ســبيل المثال، يمكن أن يختلف بمرور الوقت، وقد أقرت المقاييس التي تقيســه بهذه النقطة (ماير Mayer, 1978). على سبيل المثال، يستخدم مقياس الاكتئاب الخاص بمركز الدراسات الوبائية (رادلوف Radloff, 1977) الذي يســتخدم على نطاق واسع تنسيقاً يطلب من المستجيبين الإشارة إلى عدد المرات التي عانوا فيها حالات مزاجية مختلفة خلال الأسبوع الماضي. وبعض المقاييس، مثــل مقاييــس القلــق (e.g., Spielberger, Gorsuch, & Lushene, 1970) تم تطويرها بأشكال مختلفة وتهدف إلى تقييم الحالات العابرة (المؤقتة) نسبياً أو السمات الدائمة نسبياً (Zuckerman, 1983).

وينبغي على الباحث أن يختار إطاراً زمنياً لمقياس نشــط (إيجابي) وليس بشكل سلبي. وتُعــد النظرية بمثابة دليلٍ مهمٍ لهذه العملية. هل تشــكل الظاهرة موضع الاهتمام جانباً أساســياً ودائماً لشــخصيات الأفراد، أم أنها من المرجح أن تعتمد على الظروف المتغيرة؟ هل المقصــود من المقياس هو الكشــف عن الاختلافات الدقيقة التــي تحدث خلال فترة زمنية قصيرة (على سبيل المثال، زيادة التأثير السلبي بعد مشاهدة فيلم حزين) أو التغييرات التي قد تتطور على مدى العمر (على ســبيل المثال، السياســات المحافظــة المتصاعدة مع زيادة السن)؟

وفي الختام، ينبغي أن تعكس أشــكال وتنســيقات البنود، بما في ذلك خيارات الاستجابة والتعليمات، طبيعة المتغير الكامن محل الاهتمام والاستخدامات المستهدفة للمقياس.

الخطوة الرابعة: مراجعة تجمع البنود الأولية من قبل الخبراء:

حتى الآن، ناقشـنا الحاجة إلى تحديد وتوضيح الظاهرة محل الاهتمام تحديداً واضحاً، وتوليد وتكوين مجموعة من البنود المناسبة، واختيار شكل الاستجابة أو الردود لتلك البنود. الخطوة التالية في هذه العملية هي قيام مجموعة من الأشخاص الذين لديهم دراية ومعرفة في مجـال المحتوى بمراجعـة تجمع البنود. وتحقق هذه المراجعـة أغراضاً متعددة تتعلق بتعظيم صلاحية المحتوى للمقياس (انظر الفصل الرابع).

أولاً، قـد تؤدي مراجعة الخبراء لتجمع البنود الخـاص بك إلى تأكيد أو إبطال التعريف الخـاص بالظاهـرة. ويمكنك أن تطلب من فريق الخبراء (على سـبيل المثال، الزملاء الذين عملوا بشـكل مكثف ومسـتفيض مع المفهوم البنائي محل البحث أو الظواهر ذات الصلة) تقييم مدى ارتباط وأهمية كل بند من البنود بما تنوي قياسـه. ويكون هذا مفيداً بشـكل خـاص إذا كنت تقـوم بتطوير مقياس يتكون من مقاييس منفصلـة لقياس مفاهيم بنائية متعـددة. إذا كانت لديك عناية كبيرة عند وضعك وتطويرك للبنود، فإن الخبراء لن يجدوا صعوبـة كبيرة في تحديد البنود التي تتوافق مع المفاهيم البنائية. في الأسـاس، فإن أفكارك حـول ما يقيسـه كل بند تمثل الفرضيـة، وردود وإجابات الخبراء هـي البيانات التي تؤكد أو تنفـي الفرضية (بمعنى إنها تمثل اختبار هـذه الفرضية). حتى إذا كان الغرض من كافة البنود مخصصاً لقياس سمة واحدة أو مفهوم بنائي واحد، فإن مراجعة الخبراء تكون مفيدة. إذا قرأ الخبراء شـيئاً ما في بند لم تكن تنوي لإدراجه وتضمينه، فقد يفعل الأشـخاص الذين يكملون مقياساً نهائياً نفس الشيء.

عـادةً ما تتضمـن آليات الحصول عـلى تقييمات تتعلق بصلة البند عـلى تزويد فريق الخـبراء بتعريفك العملي للمفهوم البنائي، ثم يُطلب منهم تقييم كل بند من حيث أهميته وصلته بالنسبة للمفهوم البنائي كما قمت بتحديده وتعريفة. وقد يستلزم هذا الأمر مجرد تصنيف أهمية أو صلة كل بند على أنها عالية أو متوسطة أو منخفضة. بالإضافة إلى ذلك، قد تدعو الخبراء إلى التعليق على البنود كلٍ على حدة على النحو الذي يرونه مناسباً، وهذا يجعـل عملهم أكثر صعوبة إلى حدٍ ما، ولكنه يمكن أن يوفر معلومات ممتازة. وقد تمنحك بعـض التعليقات الثاقبة أو المتبصرة حول سـبب غموض بعض البنود، على سـبيل المثال، منظوراً جديداً عن كيفية محاولة قياس مفهومك البنائي.

كـما يمكن للمراجعين أيضاً تقييم وضوح البنود ومدى إيجازها ودقتها. قد يكون محتوى أحـد البنود ذا صلة بالمفهوم البنائي، ولكن صياغته قد تكون فيها إشـكالية. وهذا يؤثر على ثبـات البند، لأن وجود بند غامض أو غير واضح، بدرجة أكبر من بند واضح، يمكن أن يعكس عوامـل خارجة عن المتغير الكامن. في التعليـمات الخاصة التي تقدمها إلى المراجعين، اطلب منهم الإشارة إلى البنود غير الملائمة أو المربكة واقتراح صيغ بديلة لها، إذا كانوا يميلون إلى ذلك.

وتتمثل الخدمة الثالثة التي يمكن أن يقدمها خبراء المراجعة في الإشارة إلى طرق الاستفادة من الظاهرة التي فشلت في إدراجها. قد تكون هناك مقاربة كاملة قد أغفلتها وتجاهلتها. على سـبيل المثال، ربما تكون قـد قمت بإدراج العديد من البنود التي تشـير إلى المرض في مجموعـة من البنود المعنية بالمعتقدات الصحية ولكنها لم تعتبر الجروح أو الإصابات بمثابة انحـرافٍ آخـر ذي صلة عـن الصحة. من خـلال مراجعة الطرق المتنوعـة التي قمت بها لتجسـيد والتعبير بها عن الظاهرة موضع الاهتمام، يمكن أن يسـاعدك المراجعون في زيادة صلاحية المحتوى للمقياس الخاص بك إلى أقصى حد.

كلمة تحذير أخيرة بشـأن رأي الخبراء: إن القرار النهائي بقبول نصيحة الخبراء أو رفضها هو مسؤوليتك كمطور للمقياس. وفي بعض الأحيان، قد لا يفهم خبراء المحتوى مبادئ بناء المقياس، وقد يؤدي هذا إلى نصيحة سـيئة. ومن التوصيـات التي واجهها كثيرٌ من الزملاء الذين لا يتمتعون بخبرة كبيرة في مجال تطوير المقياس هي التخلص من البنود التي تتعلق بنفس الشيء. وكما سـبقت مناقشـته، فإن حذف جميع حالات التكرار من تجمع البنود أو المقياس النهائي سـيكون خطأً فادحاً لأن التكرار هو جزء لا يتجزأ من الاتسـاق الداخلي. ومع ذلك، قد يشـير هذا التعليق إلى أن صياغة البنود، ومفرداتها وتركيبها اللغوي متشابهة للغايـة ويمكن تحسـينها. انتبه جيدًا لكل الاقتراحات التـي تتلقاها من خبراء المحتوى، ثم اتخذ قرارك عن علم وبصيرة حول كيفية استخدام نصائحهم.

وفي هـذه المرحلة مـن العملية، يكون لدى مطور المقيـاس مجموعة من البنود التي تمـت مراجعتهـا من قبل الخبراء وتعديلها وفقًا لذلك. وقـد حان الوقت الآن للتقدم إلى الخطوة التالية.

الخطوة الخامسة: النظر في إدراج وتضمين البنود التي تحقق الصدق:

من الواضح أن جوهر استبيان تطوير المقياس هو مجموعة البنود التي سينبثق منها المقياس قيد التطوير. وعلى أية حال، فإن التبصر وبعض الحكمة يمكن أن تؤتي ثمارها بشكل رائع. وقد يكون من الممكن والملائم نسبياً إدراج وتضمين بعض البنود الإضافية في الاستبيان نفسه التي ستساعد في تحديد صدق المقياس النهائي. هناك نوعان على الأقل من البنود التي ينبغي مراعاتها وأخذها في الاعتبار.

النوع الأول من البنود الذي قد يختار مطور المقياس إدراجه وتضمينه في استبيان يعمل أو يساعد على اكتشاف العيوب أو المشكلات. قد لا يجيب المستجيبون على البنود ذات الأهمية الأساسية للأسباب التي تفترضها، وقد تكون هناك دوافع أخرى تؤثر على استجاباتهم وردودهم. إن العلم بذلك ومعرفته في وقت مبكر يُعد أمراً مفيداً. ومن بين أنواع الدوافع التي يمكن تقييمها بسهولة إلى حد ما هو المرغوبية الاجتماعية social desirability. وإذا كان الفرد متحمساً بقوة لتقديم نفسه أو نفسها بطريقة يعتبرها المجتمع إيجابية، فقد تكون الاستجابات على البنود مشوهة أو محرفة. ويسمح تضمين مقياس المرغوبية الاجتماعية للباحث بتقييم مدى تأثر كل بند من البنود على حدة بشدة بالمرغوبية الاجتماعية. وينبغي اعتبار البنود التي ترتبط ارتباطاً كبيراً بدرجة المرغوبية الاجتماعية التي يتم الحصول عليها مرشحة للاستبعاد ما لم يكن هناك سبب نظري سليم يشير إلى خلاف ذلك. وقد تم وضع وتطوير مقياسٍ موجزٍ ومفيدٍ للمرغوبية الاجتماعية من قبل "ستراهان وجيرباسي" (1972) Strahan and Gerbasi. ويمكن إدراج هذا المقياس المكون من عشرة بنود بشكل ملائم في استبيان.

وهناك مصادر أخرى لبنود الكشف عن اتجاهات الاستجابة غير المرغوب فيها "أناستاسي" (Anastasi, 1968). يتضمن اختبار مينيسوتا المتعدد الأوجه للشخصية The Minnesota Multiphasic Personality Inventory (هاثاواي وماكنلي، Hathaway & McKinley, 1967؛ هاثاواي وميهل، Hathaway & Meehl, 1951) العديد من المقاييس التي تهدف إلى الكشف عن أوجه التحيز المختلفة في الاستجابة. وفي بعض الحالات، قد يكون من المناسب إدراج هذه الأنواع من المقاييس.

أمــا الفئة الأخرى من البنود التــي ينبغي أخذها في الاعتبــار لإدراجها في هذه المرحلة فهي تتعلق بصدق المفهوم البنائي للمقياس. ووفقاً لما تمت مناقشــته في الفصل الرابع، إذا كانت النظرية تؤكد أن الظاهرة التي تقوم بقياســها ترتبط بمفاهيم بنائية أخرى، فإن أداء المقيــاس في مقابل مقاييس تلــك المفاهيم البنائية الأخرى يمكن أن يكون دليلاً على صدقه. وبدلاً من بذل جهد منفصل للتحقق من الصدق بعد تكوين المقياس النهائي، قد يكون من الممكــن إدراج مقاييس المفاهيم البنائية ذات الصلة في هذه المرحلة. ويمكن أن يوفر نمط العلاقات الناتجة دعماً للادعاءات المتعلقة بالصدق أو بدلاً من ذلك، يقدم أدلة إذا لم يكن أداء مجموعة البنود كما كان متوقعاً.

الخطوة السادسة: تطبيق البنود على عينة تطوير المقياس:

بعــد تحديد البنود المرتبطة بالمفهوم البنائي والصــدق التي يجب تضمينها وإدراجها في الاســتبانة الخاصة بك، يجب عليك تطبيقها، جنبــاً إلى جنب مع مجموعة البنود الجديدة، على بعض الأشــخاص. يجب أن تكون عينة الأشــخاص كبيرة. كم يبلغ حجمها الكبير؟ من الصعب التوصل إلى توافق في الآراء بشــأن هذه المســألة. دعونا ندرس الأســاس المنطقي لعينة كبيرة. ويشــير «نونالي» Nunnally (1978) إلى أن مســألة أخــذ العينات الأوَّلية في تطوير المقياس تنطوي على أخذ عينات من البنود من عالم افتراضي (راجع غيسيلي، كامبل، وزيديــك، Ghiselli, Campbell, & Zedeck, 1981). ومن أجل التركيز على مدى كفاية البنــود، ينبغي أن تكون العينة كبيرة بما فيه الكفاية للقضاء على تباين الأشــخاص باعتباره مصدر قلق كبير. ويشير «نونالي Nunnally» إلى أن ٣٠٠ شخص هم عدد كافٍ. ومع ذلك، فإن التجربة العملية تشير إلى أن المقاييس قد تم وضعها وتطويرها بنجاح باستخدام عينات أصغر حجماً. كما أن عدد البنود وعدد المقاييس التي سيتم استخراجها واستخلاصها له تأثير أيضاً على مسألة حجم العينة. وإذا كان الهدف استخراج مقياس واحد فقط من مجموعة تضم حوالي ٢٠ بنداً، فإن عينة حجمها أقل من ٣٠٠ شخص قد تكون كافية.

هناك العديد من المخاطر في اســتخدام عدد قليل جداً من الأشخاص. أولاً، قد لا تكون أنمــاط التغاير بين البنود مســتقراً. قد يتحول بند ما يبدو أنه يزيد من الاتســاق الداخلي إلى أن يكون عديم القيمة عند اســتخدامه في عينة منفصلة. إذا تم اختيار البنود لإدراجها

(كما هي بشــكل جيد للغاية) على أســاس مســاهمتها في معامل "ألفا"، فإن وجود عينة تطويرية صغيرة يمكن أن ترســم صورة وردية غير دقيقة للاتســاق الداخلي. وعندما تكون نسبة الأشخاص إلى البنود منخفضة نسبياً وحجم العينة غير كبير، يمكن أن تتأثر الارتباطات المتبادلــة بين البنود بالصدفــة chance بدرجة كبيرة إلى حدٍ مــا. وعندما تتم إعادة إدارة (تطبيــق) المقياس الذي تم اختيار بنوده في ظل هذه الظروف، فإن العوامل المحتملة التي جعلت بعض البنود تبدو جيدة في البداية لم تعد صالحة. وبالتالي، فإن معامل "ألفا" الذي يتم الحصول عليه في مناســبات «دراســات» أخرى غير دراسة التطوير الأوَّلية للمقياس قد تكــون أقــل من المتوقع. وعلى نحو مماثل، قد يتم اســتبعاد أي بند من المحتمل أن يكون جيداً لأن ارتباطه ببنود أخرى كانت منخفضة بمحض الصدفة البحتة.

وثمــة مشــكلة ثانية محتملــة لحجم العينة الصغــير هي أن عينة التطويــر قد لا تمثل المجتمع الإحصائي الذي يســتهدفه المقياس. وبطبيعة الحال، قد يكون هذا هو الحال أيضاً إذا كانــت عينــة التطوير كبيرة، ولكن من المرجح أن تســتبعد عينة صغــيرة أنواعًا معينة مــن الأفراد. وبالتالي، يجب على مطــور المقياس أن يأخذ في الاعتبار كلاً من حجم وتكوين عينــة التطوير. قد يختــار الباحث الدقيق معالجة إمكانيــة أو قابلية تعميم المقياس عبر المجتمعات الإحصائية (أو بعض الجوانب facet الأخرى) باستخدام دراسة إمكانية التعميم G-study، على النحو الذي تم توضيحه في الفصل الثالث.

ليست كل أنواع عدم تمثيل المجتمع متطابقة. هناك على الأقل طريقتان مختلفتان قد لا تمثل العينة من خلالهما المجتمع الإحصائي الأكبر. الأولى تتضمن مستوى السمة الموجودة في العينة مقابل مســتواها في المجتمع الإحصائي المســتهدف. فعلى ســبيل المثال، قد تمثل العينة نطاقًا أضيق من السمة عما كان متوقعاً في المجتمع الإحصائي. قد يكون هذا التضييق في النطاق غير متماثل أيضًا بحيث يكون متوســط الدرجــات التي يتم الحصول عليها على مقياس العينة أعلى أو أقل بكثير مما يتوقعه الفرد بالنسبة للمجتمع الإحصائي. قد تختلف الآراء المتعلقة بالعمر القانوني الملائم لشرب الخمر في حرم الجامعة، على ســبيل المثال، عن الآراء المتعلقة بالموضوع نفسه في المجتمع المحلي ككل. ولا تعني القيمة المتوسطة للسمة التــي لا تكون ممثلة للمجتمع بالضرورة عدم أهليــة العينة لأغراض تطوير المقياس. وقد يؤدي ذلك إلى توقعات غير دقيقة بشأن متوسطات المقياس في حين أنها لا تزال تقدم صورة

دقيقة عن ثبات الاتســاق الداخلي الذي يتمتع به المقياس. فعلى ســبيل المثال، قد تؤدي عينة من هذا النوع إلى استنتاجات صحيحة حول البنود الأكثر ارتباطًا مع بعضها البعض.

وهناك نوع أكثر إزعاجاً من عدم تمثيل العينة للمجتمع ينطوي على عينة تختلف نوعيًا وليــس كميــاً عن المجتمع الإحصائي المســتهدف. على وجه التحديد، فــإن العينة التي قد تختلف فيها العلاقات بين البنود أو المفاهيم البنائية عن المجتمع الإحصائي هي سبب يدعو للقلق. إذا كانت العينة غير معتادة أو اســتثنائية تمامًــا، فقد يكون للبنود معنى مختلف بالنســبة لهم عن معناها بالنســبة للأشــخاص بشــكل عام. قد تعكس أنماط الارتباط بين البنود ســمات غير عادية مشتركة بين أفراد العينة ولكنها نادرة الوجود في المجتمع الأوسع. وبعبارة أخرى، قد تكون مجموعات البنود المترابطة التي تنشأ (من التحليل العاملي، مثلاً) غير نمطية أو غير مألوفة. إذا تم ذكر ذلك بشــكل رســمي بعض الشيء، فقد يكون الهيكل أو البناء السببي الأساسي الذي يربط المتغيرات بالدرجات الحقيقية مختلفًا إذا كانت العينة تختلــف عن المجتمع الإحصائي من نواحٍ هامة. ولنتأمل هنا بعض الأمثلة الواضحة إلى حدٍ ما: إذا لم يفهم أعضاء العينة المختارة كلمة رئيســية تتكرر بين البنود وذات صلة بالمفهوم البنائي، فإن اســتجاباتهم عليها ربمــا تخبرنا بالقليل أو لا تخبرنا بــأي شيء حول كيفية أداء المقيــاس في ظــل ظروف مختلفة. فكلمــة مريض "sick" تعني "مريــض ill" في الولايات المتحــدة، ولكنها تعني "الشــعور بالغثيان nauseated" (أي الشــعور بمرض في المعدة) في إنجلــترا. قد يكون لمجموعة من الأســئلة حول المــرض illness التي تم تطويرها لمجموعة مــا معنى مختلف بشــكل ملحوظ بالنســبة للمجموعة الأخــرى. إذا كان المقياس يتعلق بمشــكلة صحية معينة لا ترتبط عادةً بالغثيان (مثل التهــاب المفاصل)، فقد تتجمع البنود التي تستخدم كلمة "مريض ill" معًا بسبب معناها المميز إذا كانت العينة بريطانية. ومن ناحيــة أخرى، فإن العينة من الولايات المتحدة الأمريكية من غير المحتمل أن تميز أو تفرق بــين العبارات المتعلقة بالمرض وغيره مــن البنود ذات الصلة بالصحة. حتى داخل الولايات المتحــدة الأمريكيــة، يمكن أن تكون لنفس الكلمة معانٍ مختلفة. فبين ســكان الجنوب في المناطــق الريفية، على ســبيل المثال، فإن تعبير "الدم الفاســد" يســتخدم أحيانًا كناية عن الأمراض التناســلية "مرض الزهري"، بينما يعني العداء في أجزاء أخرى من البلاد. وإذا كان

هناك بند يناقش "الدم الفاسد بين الأقارب" وكان أداؤه مختلفاً بين عينة من أهل الجنوب في المناطق الريفية مقابل عينات أخرى، فإن هذا لن يكون بالأمر المفاجئ أو المستغرب.

يمكن أن تؤدي عواقب هذا النوع الثاني من عدم تمثيل العينة للمجتمع إلى إلحاق ضرر بالغ بجهود تطوير المقياس. قد تكون البنية الأساسية التي تنشأ - وهي أنماط التغاير بين البنود المهمة للغاية لقضايا ثبات المقياس - بمثابة اختلاف من العينة المستخدمة في التطوير. إذا كان لدى الباحث سبب للاعتقاد بأن المعنى المنسوب للبنود قد يكون غير شائع أو غير نمطي بين عينة التطوير، فينبغي توخي الحذر الشديد في تفسير النتائج التي يتم الحصول عليها من تلك العينة.

الخطوة السابعة: تقييم البنود:

وبعد تطوير المجموعة الأولية من البنود وتدقيق هذه البنود (تطبيقها) على عينة كبيرة وممثلة بشكل مناسب، فقد حان الوقت لتقييم أداء كل بند على حدة بحيث يمكن تحديد البنود المناسبة التي تشكل أو تكون المقياس. وهذا، من جوانب عديدة، هو جوهر أو أساس عملية تطوير المقياس. وربما يأتي تقييم البند في المرتبة الثانية فيما يتعلق بتطوير البنود من حيث أهميته فقط.

الفحص المبدئي لأداء البنود:

عند مناقشة تطوير البند، أشرنا إلى بعض الصفات المرغوب فيها التي ينبغي أن تكون في بنود المقياس. دعونا نعيد النظر في هذه القضية أو المسألة. إن الجودة النهائية التي نسعى إلى وجودها في أي بند هي أن يكون ارتباطه عالٍ مع الدرجة الحقيقية للمتغير الكامن. ويأتي هذا مباشرةً من مناقشة الثبات في الفصل الثالث. ولا يمكننا أن نقيم الدرجة الحقيقية بشكل مباشر (إذا استطعنا القيام بذلك، فربما لن نحتاج إلى مقياس)، وبالتالي، لا يمكننا أن نحسب ارتباطها المتبادل مع البنود بشكل مباشر. ومع ذلك، يمكننا إجراء استدلالات بناءً على نماذج القياس الرسمية التي تمت مناقشتها حتى الآن. عند مناقشة الاختبارات المتكافئة في الفصل الثاني، أشرت إلى أن الارتباط بين أي بندين يساوي مربع الارتباط بين أي بند من البندين والدرجة الحقيقية. هذه القيمة التربيعية هي ثبات كل بند من البنود.

لذلــك؛ يمكننا التعرف على العلاقات مــع الدرجات الحقيقية من خلال الارتباطات المتبادلة بــين البنود. وكلما كانت الارتباطات بين البنــود مرتفعة، كان ثبات كل بند من البنود على حــدة مرتفعاً (أي، كلما كانت البنود مرتبطــة ارتباطاً وثيقاً بالدرجة الحقيقية). وكلما كان كل بنــد من البنود على حدة أكثر ثباتاً، كان المقيــاس الذي يؤلفونه أكثر ثباتاً (على افتراض أنها تشــترك في متغير كامن مشترك). لذلك؛ فإن أول صفة للجودة نسعى إليها في مجموعة من بنود المقياس هي أنها تكون مترابطة ترابطا قوياً. وإحدى الطرق لتحديد كيفية أو مدى الارتباط المتبادل بين البنود تتمثل في فحص مصفوفة الارتباط.

عكس تسجيل الدرجات:

إذا كانــت هناك بنود ارتباطها ارتباط ســلبي مع بنود أخــرى، فينبغي عندئذ النظر في مدى مناســبة إجراء تســجيل درجات تلك البنود في صورة عكســية (أي التصحيح العكسي لتلك البنود). وفي وقت سابق، اقترحت أن البنود التي صيغت في اتجاهات متعاكسة يمكن أن تثير مشــاكل. ولكــن في بعض الأحيان، قد ينتهي بنا الأمــر دون قصد إلى بنود مرتبطة مع بعضها البعض ارتباطاً ســلبياً. وقد يحدث هذا، على ســبيل المثال، إذا توقعنا في البداية مجموعتين منفصلتين من البنود (تتعلق بالســعادة والحزن، على سبيل المثال)، ولكننا قررنا لســبب ما أن يتم دمجهــما في مجموعة واحدة. ويمكننا بعد ذلــك أن ينتهي بنا الأمر إلى عبارات ترتبط بشــكل متســاوٍ بالمفهوم البنائي المركب الجديد (على ســبيل المثال، التأثير)، ولكــن بعضها قــد يكون إيجابياً وبعضها الآخر ســلبياً. «أنا ســعيد» و«أنا حزين» كلاهما يتعلــق على حد ســواء بالتأثير؛ ومع ذلك، فهما متضــادان. إذا أردنا الحصول على درجات عالية على مقياسنا لقياس السعادة، فسيتعين علينا أن نعطي قيمة عالية لتأييد أو الموافقة على البند «سعيد» وإعطاء قيمة منخفضة لتأييد البند «حزين». وهذا يعني، أننا سنعكس تســجيل درجة بند الحــزن. وفي بعض الأحيان، تتم إدارة أو تطبيــق البنود بطريقة تكون معكوســة بالفعل. فعلى سبيل المثال، قد يُطلب من الأشخاص وضع دائرة حول قيم رقمية أعلى لبيان الاتفاق مع بند «ســعيد» وقيم أقل للإشــارة إلى تأييــد بند «حزين». ومن بين الطرق للقيام بذلك أن تكون التوصيفات اللفظية لخيارات الاســتجابة (على ســبيل المثال، «لا أوافــق بشــدة»، «لا أوافق إلى حدٍ ما»، وما إلى ذلك) دائماً بالترتيب نفســه لكل البنود ولكن مع ربط الأرقام بها إما تصاعدياً أو تنازلياً، اعتماداً على البند:

١- أشعر بالحزن في كثير من الأحيان.

٦	٥	٤	٣	٢	١
غير موافق بشدة	غير موافق إلى حدٍ ما	غير موافق بشكل معتدل	موافق بشكل معتدل	موافق إلى حدٍ ما	موافق بشدة

٢- في كثير من الأوقات، أكون سعيداً.

١	٢	٣	٤	٥	٦
غير موافق بشدة	غير موافق إلى حدٍ ما	غير موافق بشكل معتدل	موافق بشكل معتدل	موافق إلى حدٍ ما	موافق بشدة

وقد تتسبب هذه العملية في إرباك الشخص المجيب. فقد يتجاهل الناس الكلمات بعد أن يدركوا أنها هي واحدة لكل البنود. ومع ذلك، ربما يكون من الأفضل تغيير ترتيب التوصيفات (على سبيل المثال، من "غير موافق بشدة" إلى "موافق بشدة" من اليسار إلى اليمين بالنسبة لبعض البنود وعكس اتجاه البنود الأخرى). وثمة خيار آخر يتمثل في أن تكون كل من الأوصاف اللفظية والأرقام المناظرة هي نفسها بالنسبة لكل البنود، ولكن يتم إدخال قيم مختلفة لبنود معينة عند ترميز البيانات. إن تغيير الدرجات لبنود معينة عند الترميز يكون أمراً مملاً ومن المحتمل أن يكون عرضة للأخطاء على حد سواء. لكل شخص، يجب أن يُعطَى كل بند يتم تسجيل درجته بطريقة عكسية، الاهتمام الخاص الذي ينطوي عليه تسجيل الدرجات العكسية، وهذا من شأنه أن يخلق العديد من الفرص للأخطاء.

وأسهل طريقة لتسجيل الدرجات العكسية هي القيام بذلك إلكترونيًا بمجرد إدخال البيانات إلى جهاز الحاسب الآلي. تستطيع بعض بيانات الحاسب الآلي التعامل مع كافة الدرجات العكسية لبيانات كل الأشخاص. إذا كانت خيارات الاستجابة تحتوي على قيم رقمية وكان التحويل المطلوب هو عكس ترتيب هذه القيم، فيمكن استخدام صيغة (معادلة) بسيطة. على سبيل المثال، افترض أنه تم تسجيل مجموعة من بنود الحالة المزاجية التي تمت صياغتها أو تنسيقها باستخدام مقياس «ليكرت» من القيمة (١) إلى القيمة (٧)، والأرقام الأعلى تشير إلى الموافقة. افترض كذلك أنه، لتسهيل الفهم، تم استخدام

نفس شكل أو صيغة الاستجابة لكل من بنود الحالة المزاجية الإيجابية وبنود الحالة المزاجية السـلبية. ومع ذلك، إذا تم تعيين درجات عاليـة في حالة الموافقة على، وتأييد بنود الحالة المزاجيـة الإيجابيـة، فإن المقياس يكون في الأسـاس مقياس الحالة المزاجيـة الإيجابية. إن التأييد أي الموافقة على بند الحالة المزاجية الإيجابية يجب أن يؤدي إلى قيمة عالية، وتأييد بند الحالة المزاجية السلبية يجب أن يسفر عن قيمة منخفضة، وهذا هو ما يمكن الحصول عليه إذا تم تغيير درجات الاسـتجابات بالنسبة لجميع البنود المزاجية السلبية من الدرجة (٧) إلى الدرجة (١)، وتغير الدرجة (٦) إلى الدرجة (٢)، وهكذا دواليك. ويمكن تحقيق هذا النوع من التحويل عن طريق إنشـاء درجة جديدة من الدرجة القديمة باسـتخدام الصيغة أو المعادلة التالية:

$$NEW = (J + 1) - OLD$$

حيث تشـير (NEW) و(OLD) إلى الدرجات المحولة والدرجات الأصلية على الترتيب، ويشير (J) إلى الرقم الأصلي لخيارات الاستجابة. في المثال المقدم آنفاً، فإن (J) تساوي الرقم (٧)، و(J + 1) سيكون مساوياً للقيمة (٨). وطرح الدرجة (٧) من الدرجة (٨) سيسفر عن الدرجة (١)، وطرح الدرجة (٦) سيسفر عن الدرجة (٢)، وهكذا.

قـد لا يمكـن تصحيح بعض الارتباطات السـلبية بين البنود عـن طريق عكس درجات البنود. على سبيل المثال، قد يؤدي عكس درجة بند معين إلى إزالة بعض الارتباطات السلبية ولكنهـا تؤدي إلى إنشـاء ارتباطات أخرى، ويشـير هذا عادةً إلى أن بعـض البنود لا تنتمي ببسـاطة لأنها لا ترتبط بشـكل متسـق ببنود أخرى. وينبغي التخلص من وحذف أي بند يرتبط إيجابياً مع بعض البنود ويرتبط سـلبياً مع البعض الآخر في مجموعة متجانسة إذا لم يؤدِ أي نمط من بنود الدرجات العكسية إلى إزالة وإلغاء الارتباطات السالبة.

ارتباطات البند بالمقياس:

إذا أردنـا الوصول إلى مجموعة من البنود المترابطة ببعضها البعض بدرجة كبيرة، فيجب أن يرتبط كل بند على حدة جوهرياً بمجموعة البنود المتبقية. ويمكننا فحص هذه الخاصية لكل بند عن طريق حسـاب ارتباط البند بالمقياس الخاص بـه item-scale correlation. وهنـاك نوعـان مـن ارتباط المقيـاس بالبنـد، هـما: الارتباط المصحـح للبنـد بالمقياس

corrected item-scale correlation والــذي يربــط بين البند الــذي يجري تقييمه بكل بنــود المقيــاس، باســتثناء البند نفســه، في حين أن الارتبــاط غير المصحح للبنــد بالمقياس uncorrected item-scale correlation يربــط البنــد المعني بمجموعة البنود المرشــحة بأكملها، بما في ذلك البند نفســه. وإذا كانت هنــاك عشرة بنود يجري النظر فيها من أجل تكوين مقياس، فإن الارتباط المصحح للبند بالمقياس لأي بند من البنود العشرة سيتألف من ارتباط هذا البند بالبنود الــتسعة الأخرى. وسيتألف الارتباط غير المصحح للبند بالمقياس من ارتباط هذا البند بكل البنود الـعشرة. ومن الناحية النظرية، فإن القيمة غير المصححة تخبرنا عن مدى أو كيفية تمثيل البند في المقياس بأكمله. وهذا مشــابه، على ســبيل المثال، للربــط بين مجموعة فرعية واحدة من اختبار الذكاء (IQ) وبين الاختبار بأكمله لتحديد ما إذا كان المقياس الفرعي وكيلاً proxy مناسباً أم لا. ومع ذلك، وعلى الرغم من أن الارتباط الكلي غير المصحح بين البنود uncorrected item-total correlation منطقي من الناحية المفاهيميــة أو النظريــة، فإن الواقع هو أن إدراج البند في المقياس من الممكن أن يؤدي إلى تضخم معامل الارتباط. وكلما قل عدد البنود في المجموعة، ازداد الفارق الذي سيحدثه إدراج أو اســتبعاد البند قيد الفحص والتدقيق. وبصفة عامة، ربما يكون من المناســب أو المفضل فحص الارتباط المصحح بين البند ومجمــوع البنود corrected item-total correlation. ويعتــبر البند ذو القيمة العالية لهــذا الارتباط المتبادل أكثر جاذبية (مرغوباً أكثر) من البند ذي القيمة المنخفضة.

تباينات البنود:

هنــاك خاصية أخــرى ذات قيمة لبند المقياس وهي التباين العالي نســبياً. لنأخذ حالة متطرفــة، إذا قام جميع الأفراد بالإجابة على بند معين بشــكل متطابق أو مماثل، فلن يميز المقيــاس على الإطلاق بين الأفراد ذوي المســتويات المختلفة من المفهــوم البنائي الذي يتم قياســه وســيكون تباينه مســاوياً للقيمة «صفر». وعلى النقيض من ذلك، إذا كانت عينة تطوير المقياس متنوعة فيما يتعلق بالسمة محل الاهتمام، فينبغي أن يكون مدى الدرجات التي يتم الحصول عليها لأحد البنود متنوعًا أيضًا. وهذا يعني وجود تباين كبير إلى حد ما. وبطبيعة الحال، لا يكون من المرغوب فيه زيادة التباين بإضافة مكون الخطأ.

وقد تكـون مقارنة تباينات البنود مفيدة أيضًا، خاصـةً إذا كان الهدف هو تطوير أداة تلبـي وتفي بافتراضات معادلة التكافؤ الأسـاسي ل «تـاو» essential tau equivalence. وعلى الرغم من أن معادلة التكافؤ الأسـاسي لـ "تاو" لا تفترض وجود تباينات متساوية بين البنود، إلا أنها تفترض وجود تغايرات متساوية بين البنود التي لها درجة حقيقية. وبالتالي، عندمـا يتم الوفـاء بافتراضات هذا النموذج، فإن الاختلافات أو الفـروق في تباينات البنود تكون نتيجة إلى عدم الاتسـاق في تباينات أخطـاء البنود. وكاعتبار عملي، قد يكون الفرق في تباينـات البنود بشـكل ملحوظ غير مرغوب فيها، لأنها قد تشـير إمـا إلى مصادر خطأ مختلفـة اختلافًا جوهريًا من بند إلى آخر أو فشـل في تلبية شرط وجود تباينات مشـتركة «تغايـرات» متسـاوية بين البند والدرجـة الحقيقة والتي يتطلبها نموذج التكافؤ الأسـاسي لـ «تاو». وبطبيعة الحال، لن ينطبق هذا إذا كان النموذج التوجيهي للتطوير هو النموذج المتجانس congeneric model، والذي لا يتطلب أي شيء غير البنود التي تشترك في متغير كامن مشترك.

متوسطات البنود:

من المرغوب فيه أيضًا أن يكون المتوسط قريباً من مركز المدى للدرجات المحتملة. على سـبيل المثال، إذا كانت خيارات الاسـتجابة لكل بند تتراوح بين الدرجة "١" (المناظرة لخيار الاسـتجابة "لا أوافق بشـدة") إلى الدرجة "٧" (المناظرة لخيار الاستجابة "أوافق بشدة")، فإن البند الذي يكون متوسطه قريباً من الدرجة "٤" سيكون بنداً مثاليًا. وإذا كان المتوسط قريباً من أحد طرفي المدى للمقياس، فقد يفشـل البند في اكتشـاف قيم معينة من المفهوم البنائي. قد يشير تجمع أو تراكم الدرجات عند القيمة "٧"، على سبيل المثال، إلى أن البند لم يكن مصاغاً بقوة كافية (أي إنه من النادر أن تجد أي شخص قد لا يوافق عليه).

بشـكل عام، فإن البنود ذات المتوسـطات القريبة جدًا من الحد الأقصى لمدى الاستجابة سـيكون لها تباينات منخفضة، والبنود التي تختلف على نطاق ضيق سـوف ترتبط ارتباطًا ضعيفاً أو سـيئاً بالبنود الأخرى. وكما ذكرنا سـابقًا، لا يمكن أن يكـون البند الذي لا يتغير vary أن يتغاير covary. وبالتالي، سـوف يميل المتوسـط غير المتوازن أو التباين المنخفض لأي سـبب من الأسباب إلى الحد من وتقليل ارتباط أي بند بالبنود الأخرى. ونتيجة لذلك،

يمكنك عادةً التركيز بشكل أساسي على نمط الارتباطات بين البنود كمقياس لقيمتها المحتملة. ومع ذلك، يُعد فحص المتوسطات والتباينات بمثابة فحصٍ مزدوجٍ مفيدٍ بمجرد إجراء اختيار مبدئي للبنود على أساس الارتباطات المتبادلة.

البعدية (الأبعاد):

إن مجموعة من البنود ليســت بالضرورة أن تشــكل مقياساً. قد لا يكون للبنود متغير أســاسي مشــترك (كما هو الحال في المؤشر index أو المتغير الناشــئ) أو قد يكون لها عدة متغــيرات. يُعد تحديد طبيعة المتغيرات الكامنة التــي تقوم عليها مجموعة بنود أمراً بالغ الأهميــة. على ســبيل المثال، الافتراض الذي يســتند عليه معامل "ألفــا" هو أن مجموعة البنود تكون أحادية البعد. وأفضل وسيلة لتحديد أي مجموعات البنود، إن وجدت، تشكل مجموعــة أحادية البعد هو التحليل العاملي. وهذا الموضــوع مهم بما فيه الكفاية بحيث يســتحق فصلاً كاملاً (انظر الفصل الســادس). وعلى الرغم من أن التحليل العاملي يتطلب أحجام عينات كبيرة، فإن تطوير المقياس بشكل عام يتطلب أيضاً عينات ذات أحجام كبيرة. وإذا كان هناك عدد قليل جداً من المجيبين على التحليل العاملي، فإن عملية تطوير المقياس برمتها قد تتعرض للخطر. وبالتالي، ينبغي أن يكون التحليل العاملي من أي نوع بشكل عام جزءاً من عملية وضع وتطوير المقياس في هذه المرحلة.

الثبات (الموثوقية):

مــن بين المؤشرات الأكثر أهمية لجودة المقياس هــو معامل الثبات، ألفا. في الواقع، إن كل مشــاكل البنود التي تمت مناقشــتها حتى الآن - تتمثل في المتوسط اللامركزي، والتباين الضعيــف، والارتباطات الســلبية بين البنــود، والارتباطات المنخفضة بــين المقياس والبند، والارتباطــات الضعيفة بين البنود وبعضها البعض - ســتميل إلى تقليل معامل "ألفا" وربما تبرر استخدام مؤشر بديل، مثل معامل "أوميغا". لذلك؛ بعد أن نختار البنود الخاصة بنا - نقــوم بالتخلص من البنود الفقيرة والاحتفاظ بالبنــود الجيدة - ويُعد معامل "ألفا" إحدى طرق تقييم مدى نجاحنا، على افتراض أن البنود تلبي وتفي بالافتراضات الخاصة باستخدام معامل "ألفا". إن معامل "ألفا" هو مؤشر على نسبة التباين في درجات المقياس التي تُعزى

إلى الدرجة الحقيقية. وهناك العديد من الخيارات لحساب معامل "ألفا"، تختلف في درجة الأتمتة (أي اســتخدام آلات الحوسبة). تحتوي بعض حزم الحاسب الآلي على برامج لتحليل البنود التي تقوم بحساب معامل "ألفا". في الحزمة الإحصائية للعلوم الاجتماعية والمعروف اختصاراً بـبرنامج SPSS، يحسب الإجراء RELIABILITY معامل "ألفا" للمقياس بالكامل ولجميع النسخ (k - 1) (أي لكل نسخة ممكنة مع حذف بند واحد). ويوفر البرنامج أيضًا الارتباطــات المصححة وغير المصححة بين المقيــاس والبند. ويتضمن برنامج "نظام التحليل الإحصائي statistical analysis system" والمعروف اختصاراً ببرنامج SAS حسابات معامل "ألفــا" كخاصية لإجــراء الارتباط، والمعروفة بـ PROC CORR. مــن خلال إدراج الخيار "ألفــا" ALPHA في بيان PROC CORR، ســيتم التعامل مــع المتغيرات المدرجة في بيان VAR على أنها مقياس وســيتم حساب معامل "ألفا" للمجموعة الكاملة من البنود وكذلك جميع مجموعات البنود (k - 1) الممكنة. كما يوفر أيضاً الارتباطات بين المقياس والبند.

وهناك خيار آخر لحســاب معامل "ألفا" هو القيام بذلك يدوياً. في حالة توفر تباينات لكل بند مــن البنود على حدة والمقياس ككل، فيمكن التعويض بهم في المعادلة أو الصيغة الأولى لـ حســاب معامل "ألفا" التي تمت مناقشــتها في الفصل الثالث. أو يمكن اســتخدام صيغة أو معادلة "سبيرمان - براون" Spearman-Brown، والتي تم تقديمها أيضًا في الفصل الثالث. وتستخدم هذه المعادلة المعلومات المتوفرة من مصفوفة ارتباط بدلاً من التباينات كأساس لحساب معامل "ألفا". ويتمثل أحد أوجه القصور في هذا النهج أو الطريقة في أن الارتباطات عبارة عن تغايرات موحدة (أي تمت معايرتها)، وقد تؤثر معايرة البنود كل على حدة على قيمة معامل "ألفا". وإذا تم الالتزام بدقة بنموذج الاختبارات المتكافئة، فهذا أمر غير منطقي أو غير مهم لأن الارتباطات يفترض أن تكون متســاوية. ومع ذلك، فهي تقريباً لا تكون متســاوية تماماً. لا يتطلب نموذج الصورة الأساســية لمعادلة "تاو" وجود ارتباطات متســاوية بين البنود، بل يتطلب فقط وجود تغايرات متســاوية. وبالتالي، فإن نسبة تباين كل بند على حدة والتي تُعزى إلى الخطأ لا تختلف في إطار هذا النموذج. ومع ذلك، ونظرًا لأن صيغة "ســبيرمان - براون" تعمل بالفعل مع متوســط الارتباطات المتبادلة بين البنود، وأن من بين الآثار المترتبة على نموذج الصورة الأساســية لمعادلة "تاو" أن متوسط ارتباطات المقياس بالبند يكون متســاوياً لكل بند، فإنه لا توجد هناك مشــكلة. ومع ذلك، يمكن أن

تكــون هناك فروق صغيرة (ولكن كبيرة في بعض الأحيــان) بين قيم معامل "ألفا" التي يتم الحصول عليها من الطرق الحسابية المستندة إلى التغاير مقابل الطرق الحسابية القائمة على الارتباط. ونظرًا لأن مصفوفة التغاير تســتخدم البيانات بشكل أنقى أو خام (بدون معايرة أي توحيد القياس)، فهي مفضلة ويجب استخدامها بشكل عام.

ومــن الناحية النظريــة، يمكن أن يأخذ معامــل "ألفا" القيم من "صفــر" إلى "الواحد الصحيــح"، عــلى الرغم من أنه من غير المرجح أن يحقق معامــل "ألفا" أيًا من هذه القيم المتطرفة. إذا كانت قيمة معامل "ألفا" سالبة، فهناك خطأ ما. وتتمثل المشكلة المحتملة في الارتباطات السالبة (أو التغايرات) بين البنود. إذا حدث ذلك، فحاول عكس الدرجات عند تســجيلها reverse scoring أو حذف البنود كما هو موضح سابقًا في هذا الفصل. ويقترح "نونــالي" Nunnally (1978) القيمة (٠,٧٠) كحد أدنى مقبول لـــمعامل "ألفا". ليس من غــير المألوف رؤية المقاييس المنشــورة ذات قيم منخفضة لمعامل "ألفــا". مختلف علماء المنهجية والباحثون المختلفون بدأوا مرتبكين بشــأن مستويات مختلفة من معامل " ألفا"، وأفضل بصفة شــخصية لمدى أو قيــم معامل «ألفا» للمقاييس البحثيــة على النحو التالي: أقــل من (٠,٦٠) غير مقبول، بــين (٠,٦٠) و(٠,٦٥) غير مرغوب، بين (٠,٦٥) و(٠,٧٠) الحد الأدنى المقبول، وبين (٠,٧٠) و(٠,٨٠) جيد، وبين (٠,٨٠) و(٠,٩٠) جيد جداً، وأكبر بكثير من (٠,٩٠) ينبغي للمرء أن يفكر في أن يقلل من طول المقياس (انظر القســم التالي). وأود أن أؤكــد أن هذه المجموعات هي شــخصية وذاتية لقيم معامل «ألفا». ولا أســتطيع الدفاع عنها على أسس عقلانية بحتة. ومع ذلك، فإنها تعكس تجربتي وخبرتي ويبدو أنها تتداخل بشكل كبير مع تقييمات الباحثين الآخرين. تنطبق القيم التي اقترحتها على معاملات «ألفا» المســتقرة. وأثناء التطوير، يتم اختيار البنود، إما بشــكل مباشر أو غير مباشر، على أســاس مســاهمتها في معامــل « ألفا». وقد يُعزى بعض التغاير الظاهــري بين البنود إلى الصدفة. لذلك؛ من المستحســن خلال مرحلة التطوير أن نســعى إلى الحصول على معاملات «ألفا» أعلى قليلاً مما ترغب فيه. وإذا ما تدهورت قيم معامل «ألفا» بعض الشيء عند استخدامها في ســياق بحثي جديد، فإنها ســتظل مرتفعة بشكل مقبول إلى حدٍ ما. وكما ذُكر سابقًا، إذا كانت عينة التطوير صغيرة، فيجب أن يكون الباحث قلقًا بشــكل خاص من أن التقديرات

الأوَّليــة لمعامــل «ألفا» التي يتم الحصول عليها أثناء تطوير المقياس قد لا تكون مســتقرة. وكما سنرى، فإن هذا هو الحال أيضًا عندما يكون عدد البنود التي تشكل المقياس صغيرًا.

ومــن المواقف أو الحــالات التي لا تنطبق عليها «نطاقات الراحــة» comfort ranges المقترحــة لمعامــل "ألفا" عندمــا يتم تطوير مقيــاس يتطلب دقة بالغــة الأهمية. وتعتبر الحالات السريرية مثالاً على ذلك. وتكون الإرشادات المقترحة «المبادئ التوجيهية» مناسبة للأدوات البحثية التي ستستخدم مع بيانات المجموعات. على سبيل المثال، من المحتمل أن يكون المقياس الذي تبلغ قيمة معامل «ألفا» له (٠,٨٥) مناســبًا تمامًا للاســتخدام في دراسة تقارن بين المجموعات فيما يتعلق بالمفهوم البنائي الذي يتم قياسه. يتطلب التقييم الفردي، ولاســيما عندما تعتمد القرارات المهمة على هذا التقييم، مســتوى أعلى من ذلك بكثير. إن المقاييس المخصصة للتشــخيص الفردي، أو التوظيــف، أو التوظيف الأكاديمي، أو غير ذلك مــن الأغراض المهمة ينبغي أن تكون لها درجة أعلى من الثبات، في منتصف التســعينيات، على سبيل المثال.

في بعــض الحالات، مثــل الحالات التي يتكون فيها المقياس من بند واحد، ســيكون من المســتحيل اســتخدام معامل "ألفا" كمــؤشر للثبات. وإذا كان ذلــك ممكنًا، فيجب إجراء تقييم معين للثبات. وقد يكون ارتباط إعادة الاختبار Test-retest correlation هو الخيار الوحيــد في حالة المقياس المكون مــن بند واحد. وعلى الرغم من أن مؤشر الثبات هذا غير كامــل، كما هو موضح في الفصل الثالث، فإنه من الواضح أنه أفضل من عدم تقييم الثبات على الإطلاق. والبديل المفضل، إن أمكن، هو تشــكيل وتكوين المقياس باستخدام أكثر من بند واحد.

وكــما هو مذكور في الفصل الثالث، فــإن معامل "أوميغا" هو بديل لمعامل "ألفا" الذي قد يكون مناسباً عندما لا يتم الوفاء بالافتراضات الخاصة باختبارات معادلة التكافؤ الأساسي لــ "تــاو" essentially tau-equivalent tests. وكما ذكرت أيضًا في ذلك الفصل، إن "دان وآخريــن" Dunn et al. (2014) و"بيترز" Peters (2014) يقدمون إرشــادات لاســتخدام البيئة الإحصائية (R) لحساب معامل "أوميغا".

الخطوة الثامنة: تحقيق الحد الأمثل لطول المقياس:

تأثير طول المقياس على الثبات:

في هــذه المرحلة من عملية تطوير المقياس، يكــون لدى الباحث تجمع من البنود التي تظهــر ثبات بدرجة مقبولة. ويتأثر معامل «ألفــا» للمقياس بخاصيتين هما: مدى التغاير بين البنود وعدد البنود في المقياس. بالنســبة للبنــود التي لها ارتباطات بين المقياس والبند يساوي متوسط الارتباط بين البنود وبعضها البعض (أي، البنود التي تكون نموذجية إلى حدٍ ما)، ســيؤدي إضافة المزيد منها إلى زيادة قيمة معامل «ألفا» وحذف المزيد منها ســيقلل قيمة معامل «ألفا». وبوجهٍ عام، تكون المقاييس الأقصر (القصيرة إلى حدٍ كبير) جيدة لأنها تلقــي عبئــاً أقل على المجيبين. ومن ناحية أخرى، فإن المقاييــس الأطول تكون جيدة لأنها تميل إلى أن تكون أكثر ثباتاً. ومن الواضح أن تعظيم أحد هذه الأصول من شــأنه أن يقلل من الآخر. ولذلك؛ ينبغي على مطور المقياس أن يعطي بعض التفكير في المفاضلة المثلى بين الإيجاز «الاختصار» والثبات.

إذا كان ثبات أي مقياس منخفضاً للغاية، فإن الإيجاز لا يكون فضيلة أو نقطة قوة. وقد يكون الأشخاص، في الواقع، أكثر استعداداً للإجابة على مقياس مكون من ثلاث بنود أكثر من اســتعدادهم للإجابة على مقياس مكون من عشرة بنود. ومع ذلك، إذا لم يتمكن الباحث من تعيين أي معنى للدرجات التي تم الحصول عليها من نســخة المقياس الأقصر، فلن يتم تحقيق أي شيء. وبناءً على ذلك، ينبغي أن تقتصر مسألة المقايضة بين الثبات والإيجاز على الحــالات التي يكون فيها لدى الباحث «ثباتاً احتياطيــاً». وإذا كان الأمر كذلك، في الواقع، فقد يكون من المناسب تبني مقياس أقصر مقابل التضحية بقليل من الثبات.

آثار استبعاد وإسقاط البنود السيئة:

سواء كان إسقاط واستبعاد البنود «السيئة» في الواقع يزيد أو يخفض قليلاً قيمة معامل «ألفا» فإن ذلك يعتمد على مدى ضعف البنود التي سيتم إسقاطها واستبعادها وعلى عدد البنــود الموجودة في المقياس. لنتأمل هنا التأثير الذي يســببه عدد كبير أو قليل من البنود والتــي تكــون «جيدة» بنفس القدر (أي التي لها ارتباطــات مماثلة مع نظيراتها): في حالة

وجــود عدد أقل من البنود، فإنه ســينتج عن إضافة كل بنــد أو حذفه تغير أكبر في معامل «ألفــا». إذا كان متوســط الارتباط المتبــادل بين البنود وبعضها البعــض لعدد أربعة بنود هــو (٠,٥٠)، فإن معامل "ألفا" ســوف يســاوي (٠,٨٠). وإذا كان هنــاك ثلاثة بنود فقط بمتوســط ارتباط متبادل بين بعضها البعض قدره (٠,٥٠)، فإن معامل "ألفا" سوف ينخفض إلى (٠,٧٥). وفي حالة وجود خمســة بنود لها نفس متوسط الارتباط فإن معامل "ألفا" لها سيبلغ (٠,٨٣). وبالنسبة للمقاييس التي تتألف من ٩ و١٠ و١١ بنداً مع متوسط ارتباطات متبادلــة بين البنود وبعضها البعض قدره (٠,٥٠)، فإن معامل "ألفا" ســيبلغ (٠,٩٠، ٠,٩١، ٠,٩٢) على التوالي. في الحالات الأخيرة، فإن معاملات "ألفا" ليســت فقط أعلى ولكن أيضًا أقرب بكثير في القيمة من بعضها البعض.

إذا كان للبنــد ارتباط أقل من المتوســط بدرجة كافية مع البنود الأخرى، فإن إســقاطه واســتبعاده سيرفع قيمة معامل "ألفا". وإذا كان متوسط ارتباطه بالبنود الأخرى أقل قليلاً (أو يســاوي أو يزيد) من المتوســط الكلي، فإن الاحتفاظ بالبند سيزيد قيمة معامل "ألفا". وقد ذكرت أعلاه أن مقياس مكون من أربعة بنود سيحصل على معامل "ألفا" قدره (٠,٨٠)، مع وجود متوسط ارتباط متبادل بين البنود وبعضها البعض قدره (٠,٥٠). ولكن ما مقدار انخفاض متوسط ارتباط أحد البنود بالبنود الثلاثة الأخرى للمساعدة في اتخاذ قرار باستبعاد هذا العنصر بدلاً من إلحاق الضرر بمعامل «ألفا»؟ أولاً، فكر في متوسط الارتباط بين البنود الذي ينبغي أن يكون عليه لمقياس مكون من ثلاثة بنود لتحقيق معامل "ألفا" قدره (٠,٨٠). ســيلزم أن يكون متوسط الارتباط بين البنود هو (٠,٥٧). لذلك، فبعد استبعاد أسوأ البنود الأربعة، ســتحتاج البنود الثلاثة المتبقية إلى متوسط ارتباط متبادل فيما بينها قدره (٠,٥٧). حتى تحافظ على قيمة معامل "ألفا" تساوي (٠,٨٠). وسيكون لثلاثة بنود متوسط الارتباط فيما بينها أقل من (٠,٥٧) معامل "ألفا" أقل من أربعة بنود يبلغ متوسط الارتباط المتبادل بــين بعضها البعض (٠,٥٠). وإذا افترضنا أن أفضــل ثلاثة بنود في مقياس مكون من أربعة بنود كان متوسط الارتباط فيما بينها يساوي (٠,٥٧)، فإن متوسط الارتباط بين البند المتبقي (وبالتالي الأســوأ) والبنود الثلاثة الأخرى لابد أن يكون أقل من (٠,٤٣) لكي يؤدي استبعاده إلى زيــادة قيمة معامل "ألفا" فعلياً. [إن وجود ثلاثة بنود يبلغ متوســط الارتباط المتبادل فيما بينها (٠,٥٧) وبند واحد يبلغ متوســط ارتباطه مع البنود الثلاثة الأخرى (٠,٤٣) ينتج

عن ذلك متوسط ارتباط متبادل بين البنود الأربعة بشكل عام قدره (٠,٥٠)]. بالنسبة لأي قيمة أكبر من (٠,٤٣)، فإن وجود بند رابع يكون أكثر فائدة من خفض متوسط الارتباط بين البنود يؤدي إلى الضرر. وبالتالي، يجب أن يكون البند «السيئ» أسوأ قليلاً من البنود الثلاثة الأخرى (٠,٥٧ – ٠,٤٣ = ٠,١٤) التي تستحق استبعادها.

والآن، فكــر في الموقف عندما يكــون هناك مقياس مكون من عــشرة بنود وله معامل «ألفا» قدره (٠,٨٠). أولاً وقبل كل شيء، يجب أن يكون متوسط الارتباط بين البنود حوالي (٠,٢٩) فقــط، مما يوضــح الطريقة التي يعــوض بها وجود المزيد من البنــود الارتباطات الأضعف فيما بينها. ولكي يحقق مقياس مكون من تســعة بنود نفس قيمة معامل «ألفا»، يجب أن يكون متوســط الارتباط بين البنود حوالي (٠,٣١). ولا بد أن يكون للبند «السيئ» متوسط ارتباط متبادل مع البنود التسعة المتبقية (٠,٢٠) أو أقل حتى يتسنى إدراجه كبند عاشر لســحب وخفض المتوســط الكلي للارتباط المتبادل بين البنود إلى ما دون (٠,٢٩). قد يؤدي الفشــل في خفض المتوســط إلى ما دون هذه القيمة إلى اســتفادة معامل "ألفا" من إدراج وتضمين البند. ويبلغ متوســط فرق الارتباط بين البنود التسعة "الجيدة" وبين البند الوحيد "السيئ" في هذه الحالة (٠,٣١ – ٠,٢٠ = ٠,١١)، وهو فرق أصغر من فرق الموجود في مثال المقياس المكون من البنود الأربعة.

ضبط وإصلاح طول المقياس:

كيــف يمكن للمــرء أن يقوم بإصلاح أو «ترقيع» tinkering طول المقياس في الممارســة العملية؟ من الواضح أن البنود التي تســهم بأقل قدر في الاتســاق الداخلي العام ينبغي أن تكــون أول البنود التي يتم النظر في اســتبعادها. ويمكن تحديد هــذه البنود بعدة طرق. يُظهر إجراء الثبات باستخدام برنامج " SPSS" وخيار "ألفا" "ALPHA" الخاص بالبرتوكول PROC CORR في برنامــج " SAS" تأثير حذف كل بند على معامل "ألفا" الكلي. وعادةً ما يكون البند الذي يكون لإسقاطه وحذفه أقل تأثير سلبي أو أكبر تأثير إيجابي على معامل «ألفا» هو أفضل بند ينبغي إسقاطه وحذفه أولاً. ويمكن أيضًا استخدام ارتباطات المقياس بالبنــد كمقيــاس للبنود القابلــة للاســتنفاذ (أي القابلة للحذف والإســقاط). ويجب أولاً إســقاط واســتبعاد تلك البنود التي لديها أدنى مستوى من الارتباطات بين المقياس والبنود.

كما يوفر البرنامج الإحصـائي "SPSS" أيضًا الارتباط المتعدد التربيعي squared multiple correlation لكل بند، يتم الحصول عليه عن طريق انحدار البند على جميع البنود المتبقية. وهــذا يكون تقدير لتشـاركية communality البند، ومدى تقاسـمه التبايـن مع البنود الأخرى. وكما هو الحال مع ارتباطات المقياس بالبند، فإن البنود التي لها ارتباطات متعددة تربيعية في أدنى مسـتوى ممكن لابد أن تكون هي البنود الرئيسـية المرشـحة للاسـتبعاد. وبصفة عامـة، تتقارب وتتلاقى هذه المؤشرات المختلفة لجـودة البند. وعادةً ما يصاحب الارتباط الضعيف للمقياس بالبند ارتباط متعدد تربيعي منخفض وانخفاض صغير، أو حتى زيادة، في معامل "ألفا" عند اسـقاط واسـتبعاد البند. يؤثر طول المقياس على دقة معامل "ألفا". ومن الناحية العملية، فإن معامل "ألفا" المحسـوب هو تقدير للثبات يعتمد على مدى ملاءمة افتراضات القياس للبيانات الفعلية. وقد لوحظ بالفعل أن معامل "ألفا" يزداد عندمـا يتم تضمين وإدراج المزيد من البنود (ما لم تكن بنود فقيرة نسـبيًا). وبالإضافة إلى ذلك، يزداد ثبات «ألفا» كتقدير للثبات مع زيادة عدد البنود. وهذا يعني أن معامل "ألفا" المحسـوب لمقياس أطول سيكون له فترة ثقة أضيق حوله أكثر من معامل "ألفا" المحسوب لمقياس أقصر. وفي جميع التطبيقات، سيسـفر المقياس الأطول عن قيم لمعامل "ألفا" أكثر تشـابهاً مقارنةً بما سيسـفر عنه مقياس أقصر. وينبغي أن تؤخذ هذه الحقيقة في الاعتبار عند تحديد طول أو إيجاز (اختصار) المقياس أثناء عملية التطوير.

وأخـيرًا، من المهم أن نتذكر أنه يجب تضمين هامش أمان في معامل "ألفا" عند محاولة تحسـين طول المقياس. قد ينخفض معامل "ألفا" إلى حـدٍ ما عندما يتم (تطبيق) المقياس على عينة أخرى غير العينة المستخدمة في تطويره.

تقسيم العينات:

إذا كانت عينة التطوير كبيرة بما فيه الكفاية، فقد يكون من الممكن تقسيمها إلى عينتين فرعيتين. ويمكن أن تكون إحداهما بمثابة عينة التطوير الأساسـية، ويمكن استخدام الأخرى للتحقق من النتائج المتحصل عليها. لذلك، على سـبيل المثال، يمكن اسـتخدام البيانات من العينة الفرعية الأولى لحساب معامل «ألفا»، وتقييم البنود، وضبط أو إصلاح طول المقياس، والتوصل إلى نسـخة نهائية مـن المقياس تبدو مثالية. ويمكن بعد ذلك، اسـتخدام العينة

الفرعية الثانية لتكرار هذه النتائج. ولن يستند اختيار البنود التي ينبغي الاحتفاظ بها على الإطـلاق إلى العينة الفرعية الثانية. وبالتالي، فـإن معاملات «ألفا» وغيرها من الإحصاءات المحســوبة لهذه المجموعة لن تظهر تأثيرات الصدفــة chance، مثل تضخم معامل "ألفا"، الذي تمت مناقشته في وقت سابق. إذا ظلت معاملات «ألفا» ثابتة إلى حد ما عبر العينتين الفرعيتــين، فيمكنــك أن تكون أكثر ارتياحاً لافتراض أن هذه القيم غير مشــوهة عن طريق الصدفــة. وبطبيعة الحال، من المرجح أن تكون العينتين الفرعيتين أكثر تشــابهاً من عينتين مختلفتين تماماً. ومن المرجح أن تكون العينات الفرعية، والتي يتم تقسيمها بشكل عشوائي من عينة تطوير بأكملها، ممثلة للمجتمع الإحصائي؛ وعلى النقيض من ذلك، قد تكون عينة جديــدة تماماً ممثلــة لمجتمع إحصائي مختلف بعض الشيء. كــما أن فترات جمع البيانات للعينتــين الفرعيتين لا يفصل بينهــا الزمن، في حين أن عينة التطويــر وعينة منفصلة تماماً، عــلى الدوام، يتم فصل الفترات الزمنية في جمع البيانات. وعلاوة على ذلك، فإن أي شروط خاصة قد يكون تم تطبيقها على جمع البيانات لعينة فرعية واحدة يتم تطبيقها بالمثل على العينــة الفرعية الأخرى. ومن الأمثلة على الظروف الخاصــة التعرض لأفراد بحث معينين، والإعدادات المادية، ووضوح طباعة الاســتبيانات. كذلك، قد تكون العينتان الفرعيتان هما المجموعتان الوحيدتان اللذان يكملان الإجابة على بنود المقياس مع كافة البنود من التجمع الأصلي للبنود التي تم رفضها في نهاية المطاف. وإذا كانت البنود المرفوضة قد مارســت أي تأثير على الاســتجابات على بنود المقياس، فإن هذه البنود ســتكون قابلة للمقارنة بالنسبة لكلتا العينتين الفرعيتين.

وعلى الرغم من التشــابه الفريد بــين العينات الفرعية الناتجة، فــإن تكرار النتائج عن طريق تقســيم عينة التطوير يوفر معلومات ذات قيمة حول اســتقرار المقياس. وتختلف العيّنتــان الفرعيتان في جانــب رئيسي واحد: ففي حالة العينة الفرعية الأولى التي اســتند على بياناتها في اختيار البنود، توجد فرصة لخلط عوامل عدم الاستقرار والصدفة مع التغاير الموثــوق بين البنود. ولا توجد فرصة من هذا القبيل لإســناد أو عزو نتائج الصدفة بصورة منهجية «منتظمة» إلى الثبات بالنســبة للمجموعة الثانيــة لأن بياناتها لم تؤثر على اختيار البنــود. وهــذا الاختلاف المهم هو ســبب كافٍ لتقييم المعلومات التــي يمكن أن يوفرها تقسيم العينات في هذه المرحلة من تطوير المقياس.

إن الطريقة الأكثر وضوحاً لتقسـيم عينة كبيرة بالقدر الكافي هو تقسـيمها إلى نصفين. ومع ذلك، إذا كانت العينة صغيرة جداً بحيث لا يمكن أن يسفر تقسيمها عن نصفين كبيرين بشكل كاف، فإنه يمكنك تقسيمها بشكل غير متساوٍ. ويمكن استخدام العينة الفرعية الأكبر في العمليـة الأكـثر أهمية لتقييم البنود وبناء المقياس واسـتخدام العينـة الفرعية الأصغر للتحقق من صلاحية المقياس cross-validation.

تمارين:

افترض أنك تقوم بتطوير مقياس «الخوف من الثعابين» الذي يحتوي على صيغ استجابة «ليكرت» مكون من ستة خيارات، وتم تطبيقه على ٣٠٠ شخص. على الرغم من أن المزيد من البنود سـيكون أمراً مرغوبًا فيه لتطوير المقياس الفعلي، إلا أنه يمكنك إكمال الخطوات التالية لهذه التمارين:

١- قم بإنشاء تجمع من عشرة بنود على شكل مقياس «ليكرت».

٢- بالنسبة لكل بند قمت بكتابته، قم بتقدير قيم مقياس "ليكرت" التي سيتم إقرارها من قبل "الشخص العادي" (أي، لا يوجد فوبيا من الثعابين أو ساحر الثعابين).

٣- اختر بنداً من تجمع البنود تعتقد أنه قد يسـتخلص استجابة متطرفة من شخص عادي، ثم أعد صياغتها للحصول على استجابة أكثر اعتدالاً.

٤- قم بتوليد وتكوين عشرة بنود أخرى من نوع مقياس «ليكرت» لقياس مفهوم بنائي آخر غـير الخوف من الثعابين، ثم قم بمزج أو خلط هذه البنود عشـوائياً مع البنود العشرة الأصلية واطلب من بعض الأصدقاء أن يشـيروا إلى ما يعتقدون حول ما يهدف كل بند من البنود إلى قياسه.

٥- مـا هي إيجابيات وسـلبيات بناء البنود التي يكون لها عدد زوجـي مقابل البنود التي يكون لها عدد فردي من خيارات الاستجابة؟

٦- باسـتخدام إما مفهوم الخوف من الثعابين أو المفهوم البنائي الأسـاسي الذي تقوم عليه المجموعة الثانية المكونة من عشرة بنود، قم بتدوين السـلوكيات التي يمكن ملاحظتها

بشكل مباشر والتي يمكن استخدامها للتحقق من صدق مقياس يقيس ذلك البناء واشرح كيف يمكنك استخدام البيانات السلوكية للتحقق من الصدق.

٧- ما قيمة معامل "ألفا" بالنسبة للمقياس إذا كان متوسط الارتباط المتبادل بين البنود العشرة لمقياس الخوف من الثعابين يساوي (٠,٣٠)؟[١]

٨- كيف يمكنك استخدام العينات المقسمة لتقدير معامل "ألفا" للمقياس والتحقق من الصدق عبر العينات؟

ملاحظات:

١- الإجابة: ألفا = [١٠ * ٠,٣٠] / [١ + (٩ * ٠,٣٠)] = ٠,٨١

الفصل السادس
التحليل العاملي

في الفصل الثاني، عند مناقشة النماذج النظرية المختلفة التي يمكن أن تصف علاقة بنود المقياس بالمتغير الكامن، تمت الإشارة إلى النمـوذج العاملي العام. هذا النموذج لم يفترض أن متغيراً كامناً واحداً فقط هو مصدر كل التغاير covariation بين البنود. ولكن بدلاً من ذلك، يسـمح هذا النموذج لمتغيرات كامنة متعددة لأن تكون كأسباب للتباين في مجموعة من البنود.

لتوضيح كيف يمكن لأكثر من متغير كامن أن يكون أساسـاً لمجموعة من البنود، سنعيد النظر في حالة محددة، وإن كان افتراضيًا، تمت مناقشته في الفصول السابقة. يمكن تعريف العديـد من المفاهيم البنائية constructs التي تهم علماء الاجتماع والسـلوك إجرائياً عند مسـتويات متعددة معينـة. وتعتبر مفاهيم التكيـف النفسي، والتأثير، والتأثير السـلبي، والقلـق، والقلق الاجتماعـي، وقلق مواجهة الجمهور أمثلة عـلى الظواهر الهرمية. حيث يمكن لكل مفهوم من هذه المفاهيم أن يسـتوعب تلـك التي تتبعه في القائمة، مع إمكانية تطوير مقاييس عند كل مسـتوى معين. افتراضياً، يمكن للبنود ذات الصياغة المختلفة بأطر زمنيـة وخيارات اسـتجابة مختلفة تتعلق إما بمسـتوى معين أو متوسـط أو عام من هذا التتابع. ونأمل أن يقوم مطور المقياس باختيار صيغ البنود التي تتوافق مع المستوى المحدد للمتغير. عندئذ، يمكن استخدام التحليل العاملي لتقييم نجاح عملية الاختيار هذه.

لجعـل هذا المثال أكثر تحديـدًا، خذ في الاعتبار مجموعة مكونة مـن ٢٥ بنداً متعلقة بالتأثـير. وتذكر أن أحد افتراضات نظرية الاسـتجابة للبند Item Response Theory هو أن البنود التي تشـكل مقياسـاً هي أحادية البعد - أي أنها تتعلق بمفهوم بنائي واحد فقط. وبالتـالي، فإن ما يهمنا هو ما إذا كانت هذه البنود الـ ٢٥ ينبغي أن تشـكل مقياسًـا عامًا أو عدة مقاييس خاصة. هل جميع البنود الـ ٢٥ تشـكل مقياسـاً واحداً أم أنه من الأنسب أن يكـون هناك مقاييس منفصلة لحالات التأثير المختلفة، مثل الاكتئاب والنشـوة والعداء والقلـق، ومـا إلى ذلك؟ بل ربما حتى سـيكون من الأفضل تقسـيم بنود التأثـير الإيجابية

والسلبية (على سبيل المثال، "سعيد" مقابل "حزين" للاكتئاب أو "متوتر" مقابل "الهدوء" للقلق) إلى مقاييس منفصلة. كيف نعرف ما هو الأنسب للبنود محل الدراسة؟ وبالأحرى، فإن السؤال الذي يطرح نفسه هو: هل مجموعة البنود المتعلقة بالعديد من حالات التأثير ناشئة من متغير كامن واحد أو من العديد من المتغيرات الكامنة؟

محاولة الإجابة على هذه الأسئلة مع استبعاد أسلوب التحليل العاملي والاعتماد فقط على الأساليب التي تمت مناقشتها في الفصول السابقة ستكون مهمة شاقة. يمكننا حساب معامل ألفا على كامل مجموعة بنود المزاجية، علماً أن معامل ألفا يخبرنا عن كمية التباين المشترك في مجموعة من البنود. إذا كانت ألفا منخفضة، فقد نبحث عن مجموعات فرعية من البنود التي ترتبط بقوة أكبر مع بعضها البعض. على سبيل المثال، قد نشك في أن بنود التأثير الإيجابية والسلبية لا ترتبط مع بعضها البعض وأن الجمع بينها يؤدي إلى خفض ألفا. لذا يفترض أن تكون قيم ألفا لهذه المجموعات الفرعية الأكثر تجانسًا (كلها إيجابية أو سلبية التأثير) أعلى. قد نفترض بعد ذلك أنه حتى المجموعات الفرعية الأكثر تجانسًا (على سبيل المثال، فصل القلق عن الاكتئاب بالإضافة إلى الإيجابي من السلبي) يجب أن تكون قيم ألفا لها أعلى. ومع ذلك، في مرحلة ما، قد نشعر بالقلق أيضًا من أن هذه المقاييس الأكثر تحديدًا والمتجانسة سوف ترتبط ارتباطًا وثيقًا مع بعضها البعض لأنها كانت فقط مجرد جوانب مختلفة لنفس حالة التأثير، وهذا من شأنه أن يوحي بأن بنود تلك المقاييس تنتمي إلى نفس المقياس وليس إلى مقاييس منفصلة. يجب التأكيد على أن ألفا عالية نسبياً ليست ضماناً على أن جميع البنود تعكس تأثير متغير كامن واحد. فإذا كان هناك مقياس مكون من ٢٥ بنداً، حيث ١٢ بنداً تعكس في المقام الأول متغيراً كامناً واحداً والـ ١٣ بنداً الأخرى تعكس متغيراً كامناً آخر، فإن مصفوفة الارتباط لجميع البنود يجب أن يكون لها بعض القيم العالية وبعض القيم المنخفضة. لذا يفترض أن تكون الارتباطات القائمة على بندين يمثلان نفس المتغير الكامن عالية، وتلك التي تعتمد على بنود تعكس بشكل أساسي متغيرات كامنة مختلفة ينبغي أن تكون منخفضة نسبيًا. ومع ذلك، قد يكون متوسط الارتباط بين البنود مرتفعًا بما يكفي لإنتاج قيمة مقبولة لـ ألفا لمقياس مكون من ٢٥ بنداً. في الواقع، للحصول على قيمة ألفا تساوي ٠,٨٠ لمقياس بهذا الطول، يجب أن يكون متوسط الارتباط بين البنود فقط ٠,١٤. وبالتالي، مع وجود بنود تقيس نفس المفهوم

البنائي مرتبطة بمقدار ٠,٢٩ (ارتباط متوسط مقبول ولكنه ليس كبيرًا بشكل غير معتاد لبنود داخل مجموعة فرعية متجانسة) ووجود بنود بين مجموعات مرتبطة بمقدار ٠,٠٠، لا تزال إمكانية الحصول على ألفا بمقدار ٠,٨٠ تقريباً قائمة. ولكن، بالطبع، يكون ألفا مناسبًا فقط لمجموعـة بنود أحادية البعد، لذلك قيمة الثبـات reliability score (أي قيمة ألفا) هذه ستكون مضللة كمؤشر على الاتساق الداخلي أو أحادية البعد (الذي نعرفه لا ينطبق في هذا المثـال الافتراضي). إن التحليل العاملي، موضوع هذا الفصل، هو أداة تحليلية مفيدة يمكن أن تخبرنـا، بما لا يمكن لمعاملات الثبات أن تفعل، عن الخصائص المهمة للمقياس. يمكن أن يساعدنا تطبيقياً أو عملياً في تحديد عدد المفاهيم البنائية أو المتغيرات الكامنة أو العوامل وراء مجموعة بنود ما.

نظرة عامة حول التحليل العاملي:

يخدم التحليل العاملي عدة أغراض متصلة. وأحد وظائفه الرئيسـية، كما لوحظ سابقاً، هو مسـاعدة الباحث في تحديد عدد المتغيرات الكامنة وراء مجموعة من البنود. وبالتالي، في حالـة بنود التأثير البالـغ عددها ٢٥ بنداً، يمكن أن يسـاعد التحليل العاملي الباحث في تحديـد ما إذا كانت هناك حاجة إلى بناء مفهومٍ بنائي واحد واسـع أو عدة مفاهيم بنائية لتوصيف مجموعة البنود. أيضاً، يمكن أن يوفر التحليل العاملي وسـيلة لتفسير التباين بين العديد من المتغيرات الأصلية المتصلة نسـبياً (على سـبيل المثال، ٢٥ بنداً) باسـتخدام عدد قليل نسـبياً من المتغيرات المنشأة (أي العوامل). وهذا يؤدي إلى تكثيف المعلومات بحيث يمكن حسـاب التباين باستخدام عدد متغيرات أقل. على سبيل المثال، بدلاً من الحاجة إلى ٢٥ درجـة (قيمة) لوصف كيفية إجابة المجيبين على البنود، قد يكون من الممكن حسـاب عدد أقل من الدرجات (ربما حتى واحدة) بناءً على دمج البنود. الغرض الثالث من التحليل العاملي هو تحديد المحتوى الموضوعي أو معنى العوامل (أي المتغيرات الكامنة) التي تفسر التبايـن بين مجموعة أكبر من البنود. يتـم تحقيق ذلك عن طريق تحديد مجموعات من البنود تتغاير covary مع بعضها البعض وتبدو أنها تُعرف متغيرات كامنة ذات معنى. إذا، على سـبيل المثال، نشـأ عاملان من تحليل بنود التأثير البالغ عددها ٢٥ بنداً، فإن محتوى البنود الفردية التي تشـكل مجموعات العوامل هذه يمكـن أن تعطي فكرة عن المتغيرات

الكامنــة التي تمثلها العوامــل. الوظيفة الرابعة للتحليل العامــلي ترتبط بالوظائف الثلاثة المذكورة ســابقًا. يســاعد التحليل العاملي في تحديد البنود ذات الأداء الأفضل أو الأســوأ. وبالتالي، تلك البنود الفردية التي لا تنســجم مع أي من فئات البنود المشتقة عاملياً أو التي تتناسب مع أكثر من واحدة من تلك الفئات العاملية يمكن تحديدها والنظر في إزالتها.

تقدم الأقســام التالية ملخصًا مفاهيميًا للتحليل العامــلي، ويمكن للقراء الذين يريدون التعرف على المعالجة الحسابية للتحليل العاملي بشكل أعمق الرجوع إلى المراجع المخصصة لهــذا الموضــوع، مثــل (1983) Cureton أو (1983) Gorsuch أو (1976) Harman أو (1984) McDonald.

أمثلة على الطرق المماثلة للمفاهيم التحليلية للعوامل:

للحصــول على تصور بديهي بما يقوم به التحليــل العاملي، يمكننا النظر في عمليتين أقل رسمية ولكنهما متماثلتان تقريبًا مع تلك التي ربما تكون أكثر دراية بها. أول هذه العمليات يتم اســتخدامها في بعض الأحيان في إدارة الموارد البشرية لتحديد الموضوعات المشتركة بين القضايا المحددة التي تبدو متنوعة والتي قد تهم أعضاء الفريق أو زملاء العمل.

مثال رقم ١:

افترض أن شركة صغيرة جديدة تريد تحديد الخصائص التي يعتقد موظفوها أنها مهمة لزملائهــم في العمل للتحــلي بها. إنهم يعتقدون أن تحديد ومكافأة الخصائص ذات القيمة على نطاق واســع ســوف تلعب دوراً هاماً في تهيئة بيئة عمل متناغمــة وتعاونية. وتقوم الشركة باســتقطاب أخصائي موارد بشرية لمســاعدتهم. يجمع هذا الأخصائي، الذي سنطلق عليــه اســم جيــم Jim، موظفي الشركة العشرة معًــا ويوضح أنه يود منهــم أن يفكروا في خصائــص زملائهم الموظفــين الذين يعتبرونها مهمة في مجموعــة التفاعلات التي قد تكون لديهــم في العمل، بدءًا من وضع المقترحات والتقارير معًا إلى التفاعل مع العملاء المحتملين معًا إلى مشــاركة طاولة في الكافتيريا - المجموعة الكاملة من التفاعلات التي قد تكون لدى الموظفين. يقترح جيم أنه لبدء العملية، يكتب كل موظف على قصاصات ورق منفصلة أكبر عدد من الخصائص المهمة التي يمكنهم تحديدها.

بعـد عدة دقائق، يكتب خلالها الموظفون أفكارهم، يطلب «جيم» من موظف متطوع قراءة إحدى أفكاره للمجموعة. تقول أليس Alice إن إحدى الخصائص التي كتبتها كانت "مسـتعد لمشاركة الأفكار". شكرها جيم وطلب منها أن تقوم بتعليق الورقة التي بها تلك الفكـرة عـلى الحائط. موظف آخر، يدعى بيل Bill، يقرأ أحـد خصائصه: "روح الدعابة" وهي، أيضاً، مسجلة على الحائط. تستمر هذه العملية مع قيام كل موظف على حدة بذكر كل الخصائـص التي قام بتدوينهـا. وبهذه الطريقة، يحدد الموظفـون بمفردهم مجموعة متنوعـة من الخصائـص التي يعتبرونها مهمة بالنسـبة لزملاء العمل. بعـد القيام بذلك، يقومون بلصق قصاصات الورق التي تحتوي على تلك الخصائص على الحائط. من بين تلك الخصائص المذكورة ما يلي:

لديه النية لمشاركة الأفكار.	ودي.
روح الدعابة.	يمكن الاعتماد عليه.
دائماً لديه الأدوات المناسبة للوظيفة.	يهتم بالتفاصيل.
ذكي.	لديه ذاكرة حديدية.
متقن.	متحفظ.
مُجد.	يعرف الكثير من العملاء المحتملين.
يفكر بمنطق.	جاد.
تجده في المواقف الصعبة.	لديه شخصية.
يستعد للمهام.	متعلم بشكل جيد.
يترك انطباعاً جيداً لدى العملاء.	جدير بالثقة.
لا يحاول الاستحواذ على كل الثناء.	مهتم بمظهره.
يعمل بروح الفريق.	يجيد الأسلوب القصصي.
مرح.	حاد الذكاء.
لديه سيارة رائعة.	شخص لديه عقيدة.
لديه خبرة كبيرة في هذا النوع من العمل.	لديـه الاسـتعداد لسـاعات طويلـة إذا استدعى الأمر ذلك.

وتستمر هذه العملية لبعض الوقت، وسرعان ما تتم تغطية الجدار بأكثر من ٣٠ ملصقة ورقيــة، كل منهــا تذكر ســمة كان يعتقد الموظف أنهــا مهمة. ثم بعد ذلك، يســأل جيم المشاركون أي الخصائص يرون أنه من الممكن ضمها مع بعضها البعض. تشير كاثرين إلى أن «ذكي» و«حاد الذكاء» متشــابهان. يأخذ جيم الورقة التي كتبت عليها كلمة "ذكي" وينقلها إلى جانب كلمة "حاد الذكاء". ويقترح فرانك أن الخاصية "متعلم بشــكل جيد" يجب أن تكــون أيضًا جزءًا من هذه المجموعة. وتتم إضافة العديد من الخصائص الأخرى إلى نفس مجموعة العبارات. ثم تلاحظ كارلا أن "ودي" و"يترك انطباعًا جيدًا لدى العملاء" متشابهة مع بعضها البعض وتختلف عن المجموعة الأخرى من العبارات التي تم تشكيلها مسبقاً. إنها تقترح أن يتم تجميع هاتين الخاصيتين في مجموعة جديدة، ثم يضاف "مرح" أيضًا إلى هذه المجموعة الثانية. الخاصيتان "متحفظ " و"مهتم بمظهره" تشــكلان نواة المجموعة الثالثة إلا أن إحدى الموظفات تعتقد أن "متحفظ" سيكون من الأفضل ضمها إلى "يستعد للمهام" بــدلاً من "مهتم بمظهره". واســتمرت هــذه العملية حتى قام جيم والموظفون بتشــكيل عــدة مجموعــات من الخصائص. تقريبًــا، تم وضع كل خاصية موصوفــة في مجموعة ما.

ثم يطلب جيم من المشاركين إعطاء أسماء -كلمة أو عبارة وصفية قصيرة— لكل مجموعة من العبارات. مجموعات العبارات المختلفة هذه أعطيت مسميات مثل "الذكاء"، "المظهر"، "الوعي"، "الشــخصية"، "الاعتمادية"، وغيرها من المسميات الأخرى. افتراضياً، كل مجموعة من العبارات تمثل مفهومًا رئيسيًا متعلقًا بتصورات الموظفين عن خصائص بعضهم البعض.

مثال رقم ٢:

بعد عدة ســنوات، قررت الشركــة تكرار التمرين. المدراء لديهم شــك في أن الأمور قد تغــيرت بما فيه الكفاية بحيــث لا تظل كل الفئات المحددة ســلفاً ذات صلة. جيم، ميسر الموارد البشرية، غير متواجد. لذا قررت كارول، أحد المديرين التنفيذيين للشركة، أن الطريقة الأسهل للحصول على معلومات مماثلة قد تتمثل في تطوير استبيان يحتوي على بيانات مثل تلك التي توصل إليها الأشــخاص في التمرين الســابق. ســيُطلب من الموظفين الإشــارة إلى مــدى أهمية كل خاصية، باســتخدام العبارات «لا على الإطــلاق»، و«إلى حد ما» و«جداً» كخيارات اســتجابة لهم. طبقت هذه الاســتبانات على الموظفــين، الذين بلغ عددهم الآن

ما يقرب مـن ١٥٠موظفاً. وعندما جمعت كارول الاسـتبانات بعد تعبئتها، اطلعت عليها لمعرفـة الخصائص الأكثر أهمية. وكان أحد الأشـياء التي لاحظتها هـو أن هناك خصائص مختلفـة مهمة لموظفين مختلفين، ولكـن هناك بعض الخصائص المحددة صنفت بشـكل مشـابه مع بعضها البعض. على سبيل المثال، الأشـخاص الذين اعتقدوا أن الخاصية "يهتم بالتفاصيل" كانت مهمة، كان من المحتمل أن يعتقدوا أن الخاصية "يسـتعد للمهام" مهمة أيضاً، والأشخاص الذين لا يعتبرون أحد هذه العناصر مهمة عادةً لا يعتبرون الأخرى مهمة أيضاً. بحثت كارول في سـبل للاسـتفادة أكثر من كومة الاستبانات وفكرت في كيفية فهمها. تذكرت كيف أنه خلال التمرين الأصلي الذي أجري قبل عدة سـنوات باستخدام قصاصات الورق على الحائط، ظهر أن هناك مجموعات أكثر مما هو مطلوب بالفعل. اعتقدت كارول أن بعض العبارات الفردية لا قيمة لها إلى حدٍ ما، وأحيانًا إذا كان لدى الشـخص نفسه أكثر من واحدة من هذه العبارات التي لا قيمة لها، فإن فئة كاملة لا قيمة لها ستنتج. وتساءلت عما إذا كانت هناك طريقة لتحديد عدد الفئات التي سيسـتغرقها تنقيح معظم ما يعتقده الموظفون بشـأن زملائهم في العمل. وكتمرين، حاولـت البحث عن مجموعات أخرى من العناصر (الخصائص)، مثل تلك الاثنتين اللتين لاحظتهما سابقاً، والتي تميل لأن تكون متماثلة عند الموظفين. بشكل رئيسي، بحثت كارول عن مجموعات من العناصر المتشابهة لا تستند فقط إلى محتواها ولكن إلى مدى تقييم الموظفين لها أيضاً. استغرق هذا وقتًا كبيرًا، ولم تكن كارول متأكدة بالفعل من أنها حصلت على كل مجموعات العبارات المهمة، لكنها شـعرت أنها قادرة على جمع بعض الأفكار المثيرة للاهتمام بهذه الطريقة من الاستبانات.

أوجه القصور في هذه الطرق:

كلا هذيـن المثالين يشـبهان من الناحيـة النظريـة التحليل العاملي ولكـن مع بعض الاختلافـات الهامة. في كلتا الحالتين، النتيجـة هي إعادة تنظيم كمية كبيرة من المعلومات المحددة في مجموعة أكثر قابلية للإدارة من فئات أكثر عمومية ولكن ذات معنى. افتراضياً، كل من عمليات إعادة التصنيف في فئات قد نتج عنها عدد قليل من الأفكار التي اكتسبت الكثـير مما غطتـه العبارات الفردية العديـدة. ومع ذلك، كان لـكلا النهجين أوجه قصور واضحة إلى حد ما. في المثال الأول، كانت هناك سيطرة بسيطة على جودة البيانات الناتجة.

هـذا القصور ينبـع من حقيقة أن الأشـخاص الذين يطورون البنود يختلفون في مسـتوى مهاراتهم في تكوين الأفكار ذات الصلة في عبارات موجزة. بعض الأشـخاص منفتحون أكثر مـن غيرهم وقد ينتهي بهم المطـاف إلى توليد المزيد من العبارات؛ ومع ذلك، فليس الحال دائمًـا أن يكون أعضـاء المجموعة الأكثر انفتاحاً هم الأكثر بصيرة. ولهذه الأسـباب وغيرها، تؤدي هذه العملية غالبًا إلى عبارات غامضة أو غير ذات صلة أو سخيفة. ولذا قد يكون من الصعب إزالة البنود ضعيفة الصياغة أو تحسـينها. اعتمادًا على مرونة المجموعة المشاركة في التمرين، قد يكون من غير المناسـب رفض هذه العبارات دون مجادلة مؤلفيها. ونتيجة لذلـك، يتم منحهم مصداقية أكـبر بقدر العبارات الأفضل التي تعكـس المزيد من الأفكار المهمـة. عـلى الرغم من التفـاوت في الجودة من بند إلى آخر، إلا أنـه قد يتم التعامل مع الجميع بشكل متساوٍ يزيد أو ينقص.

إذا تم تقديم العديد من البنود السخيفة ولكنها متشابهة، فمن المحتمل أن تشكل فئة تسـتند فقط إلى تشـابهها. ربما يتم ترتيب الفئات حسـب الأولوية، ولكن يتم ذلك عادةً بتوافـق الآراء، واعتـمادًا على من قام بإعداد العبارات، قـد يكون هناك إحجام عن وصف بعـض الفئات بأنها تافهة. علاوة على ذلك، تشـير تجربتي مع هذا النوع من التمارين إلى وجود ميل قوي لوضع كل عبارة في فئة معينة. يبدو أن وجود العديد من الفئات الواضحة ومـن ثم وجود عبارة واحدة أو اثنتين يتامى لا تنتمي لأي من الفئات يجعل الأشـخاص لا يوقفون عملية التصنيف حتى يتم توزيع تلك العبارات على الفئات، حتى وإن كان التوزيع في لحظتهـا غير ملائـم. علاوة على ذلك، حتى عندما يمكن لشـخص ما تحديد أي عبارات محددة تجسد فئة ما، فليس من الواضح بالضرورة تحديد أيها أفضل أو أسوأ.

لذلك؛ كيف كان أداء هذا الأسـلوب بشـكل عام بالنسـبة إلى وظائف التحليل العاملي المحددة سابقاً؟ من الواضح أن هذه الطريقة جعلت من الصعب تحديد البنود ذات الأداء الأفضل أو الأسوأ. تتعارض مراقبة جودة البنود مع الطبيعة الجماعية لهذه المهمة وقد تخاطر بالإسـاءة إلى بعض المشاركين. وعلى الرغم من أن الطريقة الموصوفة للتو لها بعض الفائدة في تحديد عدد المتغيرات الكامنة التي تستند إليها العبارات التي قام الموظفون بإنشائها، مع إسـناد هذا التحديد إلى الانطباعات الذاتية للموظفين، كما هو الحال، يبدو أقل من المثالي. عـلى الرغم من أن الموظفين قـد يكون لديهم رؤى قيمة في مكان عملهم، فمن غير المرجح

أن يكون لديهم خبرة في تنظيم هذه الأفكار إلى مجموعات متماسكة تمثل مفاهيم بنائية constructs مهمة. علاوة على ذلك، هناك القليل من الأساس الموضوعي لتحديد ما إذا كان الموظفون كمجموعة قد قاموا بعمل جيد أو سيء لتحديد عدد المتغيرات الكامنة وراء العبارات المنشأة، لأن العملية تعتمد فقط على معايير ذاتية. وفيما يتعلق بتجميع المعلومات، تظهر مشكلة مماثلة. على الرغم من أنه قد يتم تجاهل بعض البنود (أو حتى الفئات بأكملها)، إلا أن المعايير المستخدمة هي بالضرورة ذاتية ولا توجد سوى وسائل قليلة للحكم بشكل موضوعي على ما إذا كانت الخيارات الجيدة قد اتخذت. الاعتبار الوحيد الذي قد لا تؤديه هذه الطريقة بشكل سيء للغاية هو تعريف المحتوى الموضوعي أو معنى العوامل. إذا كانت العوامل ذات مصداقية (وهي مفتوحة للسؤال)، فإن فحص محتوى العبارات المقدمة قد يعطي نظرة ثاقبة في المفاهيم البنائية الأساسية التي كان الموظفون يدركون أهميتها.

المثال الثاني تجنب بعض أوجه القصور هذه. يمكن لكارول التخلص من البنود التي لا ترى مناسبتها لكونها غير ذات صلة، على الرغم من أن هذا يضيف عبئاً كبيراً على حكمها. على الأقل كانت عملية تأييد البنود أكثر ديمقراطية إلى حدٍ ما. على كل شخص أن يقيم كل بند دون التعرض لخطر استبعاد زميل العمل. وبالتالي، فإن عملية تحديد البنود ذات الأداء الأفضل أو الأسوأ ربما كانت على الأقل أكثر اتساقًا. علاوة على ذلك، لم يتم تحديد المجموعات من خلال مجرد شعور بالتشابه الواضح في العبارات ولكن عن طريق الأدلة على أن الناس تفاعلوا مع البنود التي تم تصنيفها في مجموعات متشابهة بطريقة مشتركة. أي إن التشابه كان سمة البنود (مجموعات معينة من التي تبدو أنها تثير تصورات مماثلة)، وليس المستجيبون (الذين تباينوا في استجاباتهم على أي بند محدد). رؤية بند واحد في مجموعة على أنه غير مهم يعني أنه من المحتمل بشكل كبير لنفس الفرد رؤية البنود الأخرى في نفس المجموعة على أنها غير مهمة. قد يرى موظف آخر أن هذه البنود نفسها مهمة بشكل متسق. المسألة الحاسمة هي أنه، بغض النظر عن تقييم الفرد للأهمية، هناك ميل لأن تكون متسقة خلال العبارات داخل المجموعة. في الواقع، كان هذا هو الأساس الذي شكلت عليه كارول المجموعات. وبالتالي، ربما قد تم تحسين عملية تحديد عدد المتغيرات الكامنة إلى حد ما من خلال اعتماد هذه المنهجية المنقحة القائمة على الاستبيان.

بطبيعة الحال، فإن تقييم البنود عن طريق الفحص البصري لـ ١٥٠ اســتبيانًا ســيكون أمرًا شــاقًا إلى حــدٍ كبير، ومن المحتمــل أن نظام التصنيف لكارول لم يكــن أكثر الطرق الممكنة كفاءة. ومن الأســئلة التي تثار: ما مدى الاتســاق المطلوب للبنود لاعتبارها مجموعة؟ كم عــدد حالات قيام موظف واحد بإعطاء تقييــمات متباينة (أي اتفاق واختلاف في الأهمية) لبندين في نفس المجموعة المحتملة يمكن أن تقبلها كارول؟ من الصعب القول إلى أي مدى تكون هــذه الطريقة قد حققت الوظائــف المتبقية - تركيز المعلومــات وتحديد المحتوى الموضوعــي أو معنــى العوامل. يعتمد كلاهما على مدى جــودة العمل الذي كانت كارول قادرة على القيام به لإيجاد واستبعاد البنود الضعيفة، والتحديد بشكل صحيح الموضوعات المهمة، ومن ثم استخدام البنود داخل المجموعات لتفسير المفاهيم البنائية المحتملة.

باختصــار، من الواضــح أن هاتين الطريقتين تتركان الكثير مما هــو مرغوب فيه لو كان الهدف هو تحقيق الوظائف المذكورة أعلاه والتي ترتبط عادةً بالتحليل العاملي. وللإنصاف، قــد لا يكون الهدف من المنهجيات الموصوفــة (خاصة الأولى) تحقيق تلك الوظائف، ولكن يبــدو أن بعــض الأهداف المماثلة لما ينجــزه التحليل العاملي تحفز القيام بممارســات مثل التمريــن الأول الــذي تم وصفه ســابقاً. في حين أنــه تم القبول بهذه الحجــة، فإن تمرين الموظــف بالمقارنــة مع التحليل العاملي يخدم الأغراض: (أ) تقديم توضيح أكثر ســهولة لما يمكن أن يحققه التحليل العاملي و(ب) تســليط الضوء عــلى أوجه القصور المحتملة لنهج ذاتي وغير رسمي لتحديد المتغيرات الأساسية.

الوصف المفاهيمي للتحليل العاملي:

التحليــل العامــلي عبارة عن فئة مــن الإجراءات التي تنجز نفــس النوع من التصنيف مثل الأســاليب المذكورة أعلاه ولكنها تفعل ذلك وفقًا لمجموعة من العمليات الأكثر تنظيماً وتوفر معلومات أكثر وضوحًا يمكن لمحلل البيانات استخدامها في إصدار الأحكام. وكالطرق الموصوفــة آنفاً، يحدد التحليل العاملي فئات من العبــارات المتماثلة. تتمثل المهمة الأولى للمحلل في تحديد عدد الفئات التي تكفي لالتقاط الجزء الأكبر من المعلومات الموجودة في مجموعة العبارات الأصلية.

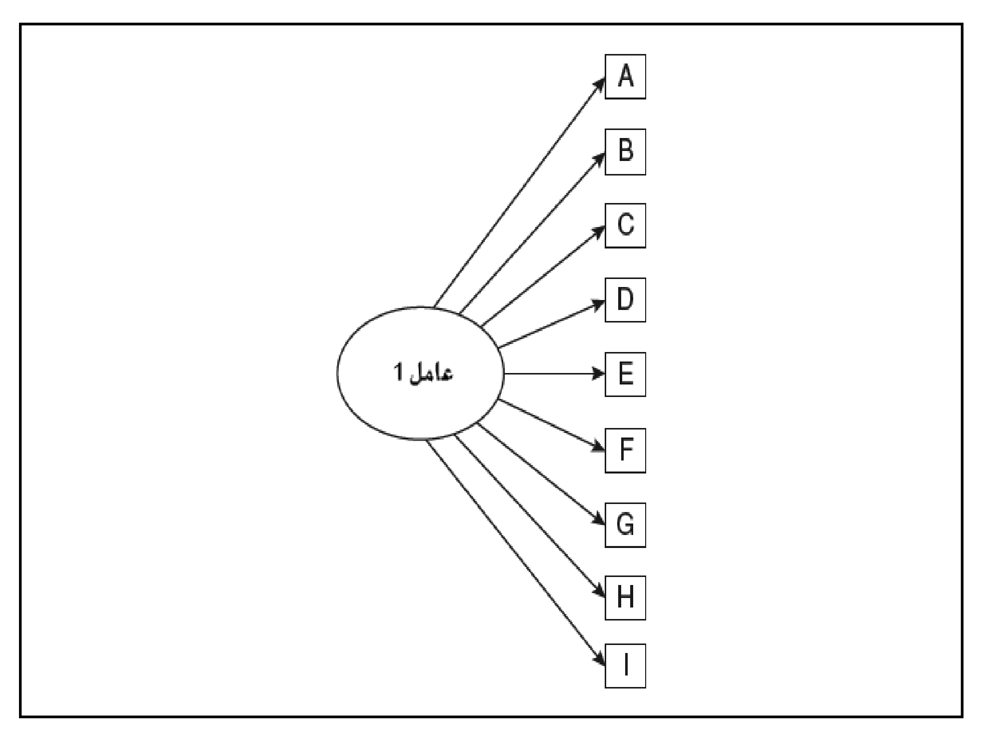
استخلاص العوامل:

يمكـــن للشـــخص أن ينظر إلى التحليل العاملي كبداية على افــتراض أن فئة واحدة كبيرة تحتوي على جميع البنود هو كل المطلوب (بمعنى، أن مفهوم أو فئة واحدة تكفي للأخذ في الاعتبار نمط الاستجابات). عندئذٍ يقوم بتقييم مقدار الارتباط بين البنود الفردية التي يمكن أن يفسرهــا ذلك المفهوم الوحيد. يقوم التحليل بعد ذلك بإجراء فحص لمعرفة مدى نجاح فرضيــة المفهوم الوحيد. إذا ظهر أن مفهومًا أو فئة واحدة لم تقم بالعمل اللازم لحســاب التغايــر covariation، بين البنود فإن التحليــل العاملي يرفض الفرضية المبدئية. ثم يحدد المفهــوم الثاني (أي المتغير الكامن أو العامل) الــذي يفسر بعض التغاير المتبقي بين البنود. تستمر هذه العملية إلى أن يكون مقدار التغاير الذي لم تفسره مجموعة العوامل صغيرًا.

العامل الأول:

كيف يتم ذلك؟ تبدأ العملية بمصفوفة الارتباط لجميع البنود الفردية. باستخدام هذه المصفوفة كنقطة بداية، يقوم التحليل العاملي بفحص أنماط التغاير الممثلة بالارتباطات بين البنود. ما يلي هو وصف مفاهيمي، وتم حذف بعض التفاصيل الرياضية لمصلحة الوضوح، لذلــك لا ينبغي أن يؤخذ ذلك حرفيًا حيث إنه فقط مجموعــة العمليات التي تقوم عليها تحليلات العوامل المولدة بالحاسوب.

كما ذكرنا سابقًا، تتضمن العملية الفرضية المبدئية للمفهوم الفردي الذي يمكن وبشكل كافٍ أن يأخذ بالاعتبار نمط الارتباطات بين البنود. وهذا بمثابة تأكيد مؤقت على أن النموذج الذي يحتوي على متغير كامن واحد (أي عامل واحد)، مع وجود مســار منفصل ينبثق منه إلى كل بند من البنود، هو تمثيل دقيق للعلاقات الســببية (انظر الشــكل رقم ٦-١). هذا يعني كذلك أن مثل هذا النموذج يمكن أن يأخذ في الاعتبار الارتباطات بين البنود. لاختبار هــذا الافتراض من الناحية المفاهيمية، يجــب أن يحدد برنامج التحليل العاملي العلاقة بين كل بند والعامل الذي يمثل المتغير الكامن الفردي، ومن ثم النظر إلى ما إذا كان يمكن إعادة إنشــاء الارتباطات المرصودة بين البنود عن طريق ضرب المســارات التي تربط كل زوج من المتغيرات عبر العامل. ولكن كيف يمكن للبرنامج حســاب الارتباطات بين استجابات البنود المرصودة والعامل الذي يمثل متغير كامن لم يتم ملاحظته أو قياسه مباشرةً؟

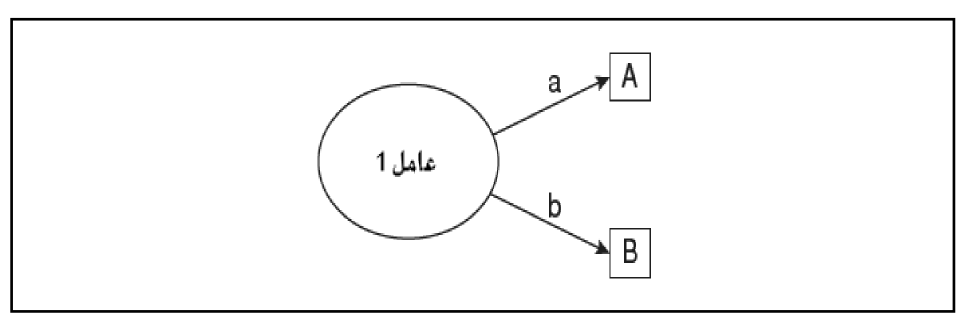


شكل (٦-١): نموذج العامل الواحد

أحد الأســاليب هو افتراض أن مجموع جميع الاســتجابات على البنود هو تقدير عددي معقــول للمتغير الكامن الشــامل الذي يُفترض أن يأخذ في الاعتبــار الارتباطات بين البنود. وبشــكل رئيسي، هــذا المجموع الكلي هو تقدير «لدرجة» المتغــير الكامن. ولأن الدرجات الفعليــة لجميع البنود يُفترض في الوقت الحالي تحديدها من خلال متغير كامن واحد، فإن كميــة المعلومــات من جميع البنــود (أي المجموع الكلي للمعلومات) هــو تقدير معقول للقيمة العددية للمتغير الكامن. من السهل إلى حدٍ ما إضافة درجات البنود معًا إلى مجموع الدرجات وحســاب ارتباطات البند - المجموع item-total correlations لكل بند منفصل مــع إجــمالي كل البنود. تعمل ارتباطــات البند - المجموع كممثــل للارتباطات بين البنود المرصودة والمتغير الكامن الغير مرصود (أي المسارات السببية من المتغير الكامن إلى البنود). مع القيم المحددة لهذه المســارات السببية، يمكن للشــخص بعد ذلك حساب الارتباط بين البنود المتوقعة بناءً على نموذج العامل الواحد. تمثل هذه الارتباطات النموذجية إسقاطات

لمـا يجـب أن تكون عليه الارتباطـات الفعلية بين البنود إذا كانت فرضية المتغير الأسـاسي الوحيـد صحيحـة. يمكن تقييم شرعية الفرضية من خلال مقارنـة الارتباطات المتوقعة مع الارتباطات الفعلية. هذا يعني طرح كل ارتباط متوقع من الارتباط الفعلي المقابل بناءً على البيانات الأصلية. يشير الفرق الكبير بين الارتباطات الفعلية والمتوقعة إلى أن نموذج العامل الواحد غير مناسب؛ ولا يزال هناك بعض التغاير بين البنود لم يتم أخذه بالاعتبار.

قد يسـاعد النظر في هذا التسلسـل لزوج واحد من البنـود، AوB، التي تعد جزءًا من مجموعـة أكبر. أولاً، سـتتم إضافة المجموعـة الكاملة من البنود، بمـا في ذلك A وB، معًا للحصول على درجة ملخصة، ثم سيتم حساب الارتباطات بين A والدرجة الإجمالية وبين B والدرجة الإجمالية. يُفترض أن هذين الارتباطين بين البند والمجموع الكلي يمثلان الارتباطات بين A وB والعامل، الذي يتوافق مع المتغير الكامن الأساسي. إذا كانت فرضية متغير كامن وحيد صحيحة، فسـيكون لرسـم تخطيطي للمسار يشـتمل على A وB والعامل مسارات (a وb في الشـكل ٦-٢) من العامل إلى كل من البندين A وB. ستكون قيم هذه المسارات هي الارتباطات بين البند والمجموع الكلي للبنود الموصوفة للتو. بناءً على مخطط المسـار هذا، يجب أن يكون الارتباط بين A وB هي حاصل ضرب هذين المسـارين. حسـاب هذا الارتباط المقترح بين A وB يتضمن الضرب البسيط.



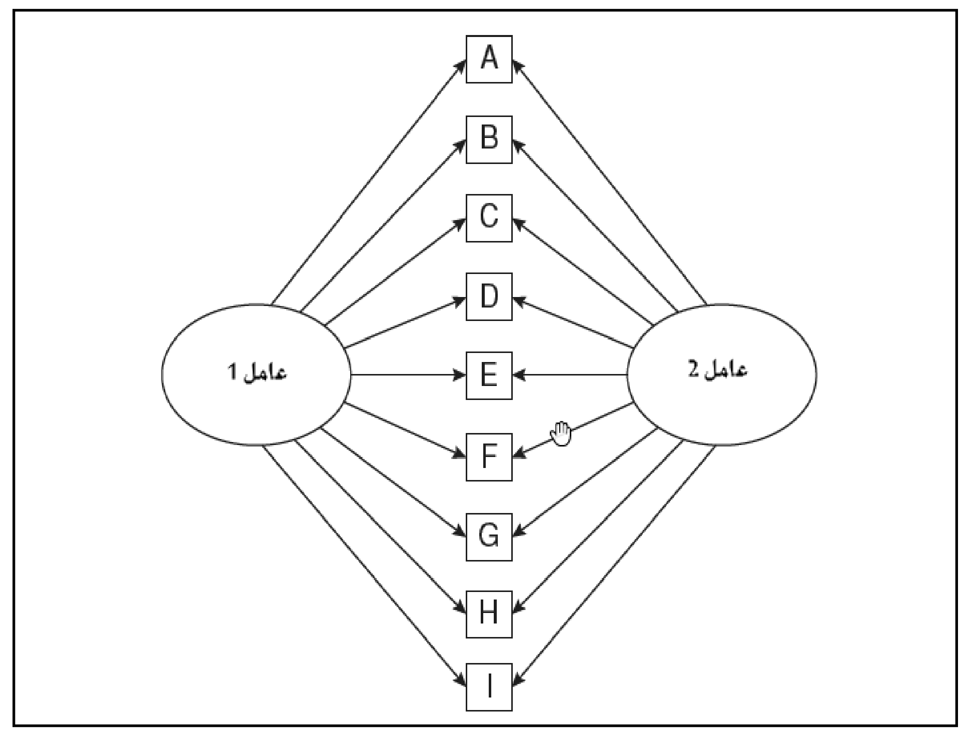
شكل (٦-٢): نموذج مبسط لعامل واحد يتضمن بندين فقط

بمجرد حسـابه، يمكن مقارنة الارتباط المقترح مع الارتباط الفعلي بين A وB. يمكن طرح الارتباط المقترح من الارتباط الفعلي للحصول على الارتباط المتبقي. قد يشير الارتباط المتبقي الكبير إلى أن الادعاء بأن متغير كامن وحيد هو السبب الوحيد للتغاير بين A وB غير كافٍ.

العمليــات التي تتم على مصفوفة الارتباط بأكملها في وقت واحد تقوم بالإجراء المذكور ســابقاً لكل زوج ممكن من البنود. وبدلاً من الانتهاء إلى ارتباط متبقٍ واحدٍ، يتم حســاب مصفوفــة كاملــة من الارتباطات المتبقية (تســمى: مصفوفة البقايــا residual matrix)، بحيــث يمثــل كل ارتباط متبقٍ مقدار التغاير بين زوج معين مــن البنود الموجودة فوق وما بعد التغاير الذي يمكن لمتغير كامن واحد أن يفسره.

العوامل اللاحقة:

أصبــح من الممكن الآن العمل على مصفوفــة البقايا هذه بنفس الطريقة التي عولجت بها مصفوفة الارتباط الأصلية، واســتخراج عامل ثانٍ مقابل لمتغير كامن جديد. مرةً أخرى، يمكن حســاب الارتباطات بين البنود والمتغير الكامن الثاني (العامل ٢ في الشــكل رقم ٦-٣)، وبناءً على تلك الارتباطات، يمكن إنشاء مصفوفة الارتباطات المقترحة.



شكل (٦-٣): نموذج بعاملين

تمثــل هذه الارتباطــات المقترحة مدى الارتباط الذي يجب أن يبقى بين البنود بعد أخذ العامل الثاني في الاعتبار. إذا اســتحوذ العامل الثاني على كل التغاير المتبقي بعد اســتخراج العامــل الأول، فينبغي أن تكون هذه القيــم المتوقعة مماثلة للقيم التي كانت موجودة في مصفوفــة البقايا المذكورة أعلاه. إذا لم يكن الأمــر كذلك، فقد تكون هناك حاجة إلى مزيد من العوامل لحساب التغاير المتبقي الذي لم يُنسب بعد إلى أحد العوامل.

يمكن أن تســتمر هذه العملية، مع كل عامل جديد مستخرج من مصفوفة البقايا التي نتجت عن التكرار الســابق، حتى يتم تحقيق مصفوفــة تحتوي فقط على علاقات متبادلة صغيرة مقبولة. عند هذه النقطة، يمكن للشــخص أن يتخذ قراراً أنه قد تم حســاب جميع التغايرات المهمة بشــكل أســاسي وأنه لا توجد عوامل أخرى مطلوبة. من الممكن متابعة العملية حتى يتم الحصول على مصفوفة بقايا تتكون بالكامل من الأصفار. وهذا سيحدث عندما يكون عدد العوامل المستخرجة يساوي عدد البنود في التحليل العاملي. وبمعنى آخر، ستتمكن دائمًا مجموعة من العوامل k من تفسير كل التغاير بين مجموعة من البنود k.

تحديد عدد العوامل التي يجب استخلاصها:

يمكن أن يكون تحديد عدد العوامل التي يمكن اســتخراجها مشكلة معقدة (على سبيل المثــال، Zwick & Velicer, 1986). وبطبيعة الحال، فــإن الدافع الرئيسي لإجراء التحليل العاملي هو الانتقال من مجموعة كبيرة من المتغيرات (البنود) إلى مجموعة أصغر (العوامل) تقــوم بعمل معقــول لالتقاط المعلومــات الأصلية (أي تركيز المعلومــات). ويمكن تحديد "العمل المعقول" بعدة طرق.

بعض الأساليب التحليلية للعوامل تستخدم معياراً إحصائياً، مثل تلك القائمة على التقدير بأسلوب الاحتمال الأعظم والتحليل العاملي التوكيدي (والذي ستتم مناقشته لاحقًا) القائم على طريقة نمذجة المعادلات البنائية. في هذا الســياق، يشــير مصطلح المعيار الإحصائي إلى حقيقة أن الأســاليب الاستدلالية تُســتخدم لتحديد ما إذا كان احتمال وجود نتيجة معينة ضئيلًا بدرجة كافية لاستبعاد حدوثها بالصدفة. هذا يعني إجراء اختبار لمعرفة ما إذا كانت البقايــا المتبقية تحتوي على مقــدار من التغاير تكون إحصائياً أكــبر من الصفر وذلك بعد استخراج كل عامل لاحق. إذا كان هذا هو الحال، تستمر العملية حتى لم يعد الأمر كذلك.

إن الاعتماد على معيار إحصائي بدلاً من الحكم الشــخصي هو ســمة جذابة لهذه الأساليب. ومع ذلك، في تطوير المقاييس، قد لا يتوافق هذا مع الهدف المطروح، والذي يتمثل في تحديد مجموعة صغيرة من العوامل التي يمكن أن تفسر التغاير المهم بين البنود. تسعى الأساليب الإحصائية إلى الحصول على تحديد شــامل للعوامــل الكامنة وراء مجموعة من البنود. إذا كان هناك بعض مصادر التغاير لم يتم حسابها من قبل أي من العوامل المستخرجة بعد، فإن مثل هذه الأساليب تستمر في استخراج المزيد من العوامل. وما يسعى إليه مطور المقياس في كثير من الأحيان بعد ذلك هو الاختزال في عدد العوامل. أي إنه في سياق تطوير المقياس، غالبًــا ما نريد معرفة مصادر الاختــلاف القليلة والأكثر نفوذاً التي تقوم عليها مجموعة من البنود، وليس كل مصدر ممكن يمكننا اكتشافه. عند تطوير المقياس، يقوم المرء عادةً بإنشاء قائمة طويلة من البنود أكثر من المتوقع للوصول إلى الأداة النهائية. والبنود التي لا تساهم في العوامل الرئيسية المحددة قد يتم حذفها في نهاية المطاف. هدفنا هو تحديد عدد قليل نســبياً من البنود التي ترتبط بقوة بعدد صغير من المتغيرات الكامنة. وعلى الرغم من أن محلــل البيانات الماهر يمكنه تحقيــق هذا الهدف من خلال الأســاليب التحليلية للعوامل باســتخدام معيار إحصائي، إلا أن الباحث الأقل خــبرة، ومن المفارقات، قد يكون أفضل من خلال استخدام إرشادات أخرى أكثر ذاتية ولكن من المحتمل أن تكون أقل غموضاً.

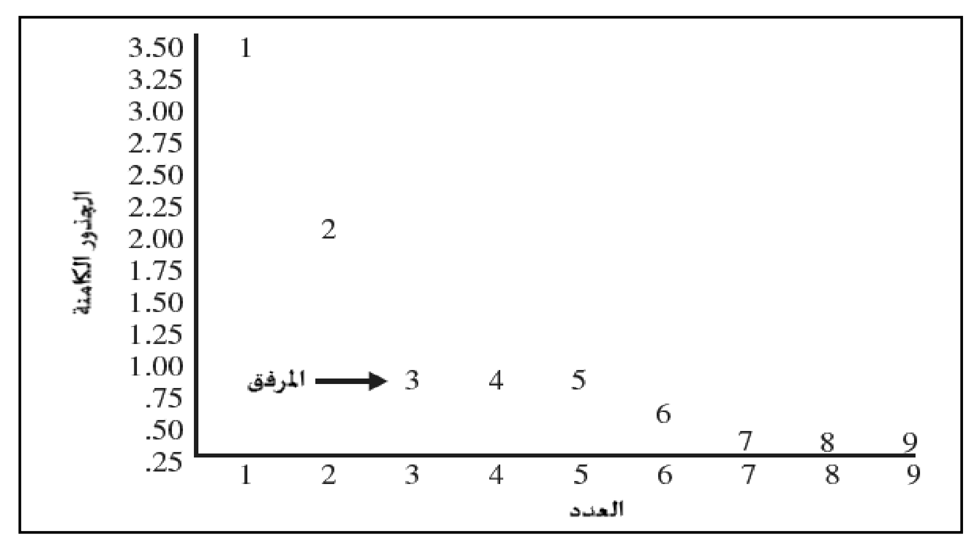
هذه الإرشــادات الذاتية تعتمد نسبيًا على نســبة التباين الكلي بين البنود الأصلية التي يمكن لمجموعة من العوامل شرحها. وهذا هو الأســاس نفســه الذي تســتخدمه الأساليب الإحصائيــة. بينــما في حالــة المعايير غــير الإحصائية (بمعنــى، لا تســتند إلى الاحتمالية)، يقــوم محلــل البيانات بتقييم مقــدار المعلومات التي يحتوي عليهــا كل عامل تالٍ ويتخذ القــرار حيال ذلك متى ما تم الوصول إلى نقطة تناقص النتائج. يشــبه هذا تقريبًا تفســير الارتباطات (مثل معاملات الثبات) على أساس حجمها، معيار ذاتي، وليس على أساس قيمتها الاحتمالية p، كمعيار إحصائي. اثنان من المبادئ التوجيهية غير الإحصائية المســتخدمة على نطاق واســع للحكم عند اســتخراج عوامل كافية هي قاعدة الجذور الكامنة eigenvalue (Kaiser, 1960) واختبار الحصاة scree test (Cattell, 1966).

يمثــل الجذر الكامن (القيمة الكامنــة) eigenvalue مقدار المعلومات التي تم التقاطها بواســطة العامل. بالنســبة لأنواع معينة من طرق تحليل العوامــل (مثل تحليل المكونات

الرئيسـية principal component analysis، والتي ستتم مناقشته لاحقًا في هذا الفصل)، فإن إجمالي كمية المعلومات في مجموعة من البنود يساوي عدد البنود. وبالتالي، في تحليل ٢٥ بنداً، سيكون هناك ٢٥ وحدة من المعلومات. الجذر الكامن لكل عامل يتوافق مع جزء مـن هذه الوحدات. على سـبيل المثال، في حالة تحليل مكـون من ٢٥ بنداً، فإن العامل ذا الجذر الكامن ٥,٠ يمثل ٢٠٪ (٥/٢٠) من إجمالي المعلومات؛ جذر كامن آخر بمقدار ٢,٥ تمثل ١٠ ٪ مـن إجمالي المعلومات، وهلم جراً. نتيجة لهـذه العلاقة بين كيفية قياس المعلومات وعـدد البنـود في التحليل هي أن الجذر الكامن بقيمـة ١,٠ يقابل 1 / k من التباين الكلي بين مجموعة من البنود. وبعبارة أخرى، فإن العامل (بافتراض تحليل المكونات الرئيسـية) الذي يحقق جذر كامن تساوي ١,٠، يحتوي على نفس النسبة من إجمالي المعلومات كما هو الحـال في البند الفردي العادي. وبالتـالي، إذا كان الهدف من التحليل العاملي هو الوصول إلى عدد أقل من المتغيرات التي تلتقط بشكل كبير المعلومات الواردة في مجموعة المتغيرات الأصليـة، فإن العوامل يجـب أن تكون محملة بالمعلومات أكثر من البنـود الأصلية. وفقًا لذلـك، تؤكد قاعدة الجذر الكامـن (Kaiser, 1960) أنه لا ينبغي الاحتفاظ بالعوامل ذات الجذور الكامنة التي أقل من ١,٠ (وبالتالي، تحتوي على معلومات أقل من البند المتوسط). على الرغم من أن الأسـاس المنطقي لاستبعاد هذه العوامل أمر منطقي، فماذا عن العوامل التي تزيد قليلاً عن ١,٠؟ هل العامل الذي يفسر معلومات أكثر بنسبة ١٪ من البند العادي حقًا يقدم نوعًا من المعلومات المركزة التي نلاحقها؟ في كثير من الأحيان، تكون الإجابة "لا"، مما يوحي بأن قاعدة الجذور الكامنة قد تكون سـخية للغاية كأساس للاحتفاظ بالعوامل. أعتقد أن هذا هو الحال عمومًا في تطوير المقياس استنادًا إلى الأساليب الكلاسيكية.

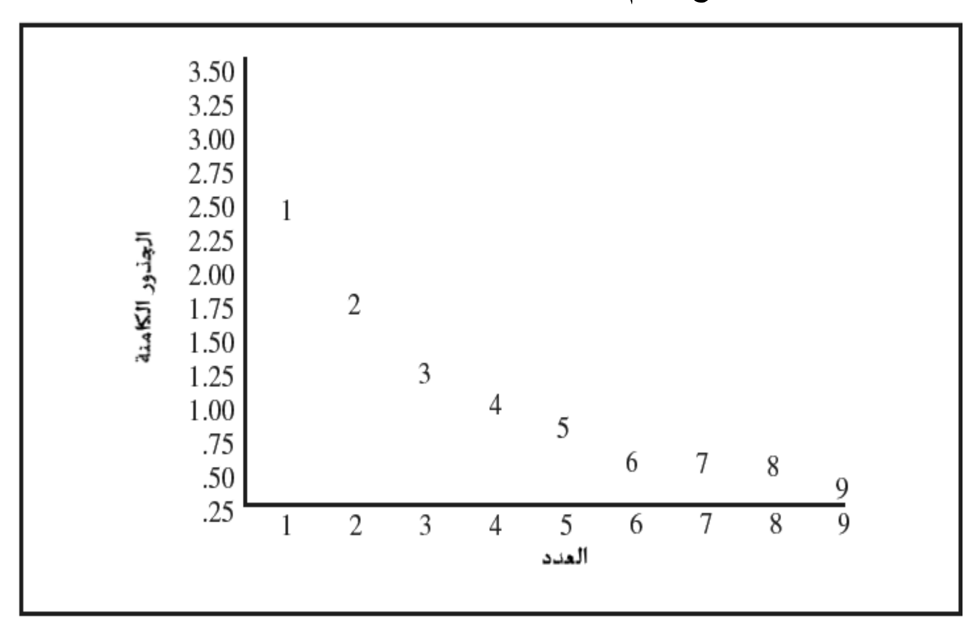
ويسـتند اختبار الحصـاة scree test (Cattell, 1966) أيضاً على الجذور الكامنة ولكن يسـتخدم قيمها النسبية بدلاً من قيمها المطلقة كمعيار. ويستند إلى الرسم البياني للجذور الكامنة المرتبطة بالعوامل المتعاقبة. ونظرًا لأن كل عامل، بعد العامل الأول، يتم استخلاصه من مصفوفة تمثل البقايا residuals الناتجة من اسـتخراج العامل السابق (كما هو موضح سابقًا)، فإن كمية المعلومات في كل عامل تالٍ تكون أقل من سابقاتها. اقترح كاتيل Cattell أنه يمكن تحديد العدد "الصحيح" من العوامل من خلال النظر في انخفاض كمية المعلومات (وبالتالي، في حجم الجذر الكامن) عبر العوامل المتعاقبة. عند رسم هذه المعلومات، سيكون

لها شــكل يتســم بجزء عمودي في الغالب على اليسار (يمثل الجذور الكامنة الكبيرة) ينتقل إلى جزء أفقي نســبيًا على اليمين (مطابق للجذور الكامنة الصغيرة). واعتبر كاتيل العوامل المرتبطة بالجانب الأيمن التي تمثل الجزء الأفقي من الرســم البياني مســتهلكة لا حاجة لها. في المصطلحــات العامة، الحصاة تصف الركام الذي يتجمع على الأرض بعد الانهيار الأرضي. يعنــي هذا المصطلح، إذن، أن الجزء الرأسي من الرســم البياني هو المــكان الذي توجد فيه العوامل الجوهرية بينما الجزء الأفقي هو الحصاة، أو الركام، التي يجب التخلص منها. من الناحية المثالية، ســيكون لتقدم العوامل نقطة تســقط عندها المعلومات فجأة، مع انتقال مفاجــئ مــن رأسي إلى أفقي و"مِرفــق elbow" واضح (انظر الشــكل ٦-٤). يدعو معيار Cattell إلى الإبقــاء عــلى تلك العوامل الكامنــة فوق المرفق من الرســم البياني. في بعض الأحيان، لا يكون الانتقال مفاجئًا لكنه تدريجي (انظر الشكل رقم ٦-٥)، مع وجود منحنى خفيف يتكون من عدة عوامل تقع بين المناطق الرأسية والأفقية للرسم البياني. في مثل هذه الحــالات، قد يكون تطبيق اختبار الحصاة لـ Cattell أمرًا خادعًا ويتطلب اعتمادًا أكبر على معايير ذاتية، مثل قابلية التفســير للعامل. يعتبر العامل قابلاً للتفســير إلى الحد الذي تبدو فيه البنود المرتبطة به متشابهة مع بعضها البعض وتجعلها ذات معنى من الناحية النظرية ومنطقية كمؤشرات لمفهوم بنائي متماسك.



شكل (٦-٤): رسم الحصاة scree plot يتضمن مرفق elbow واضح

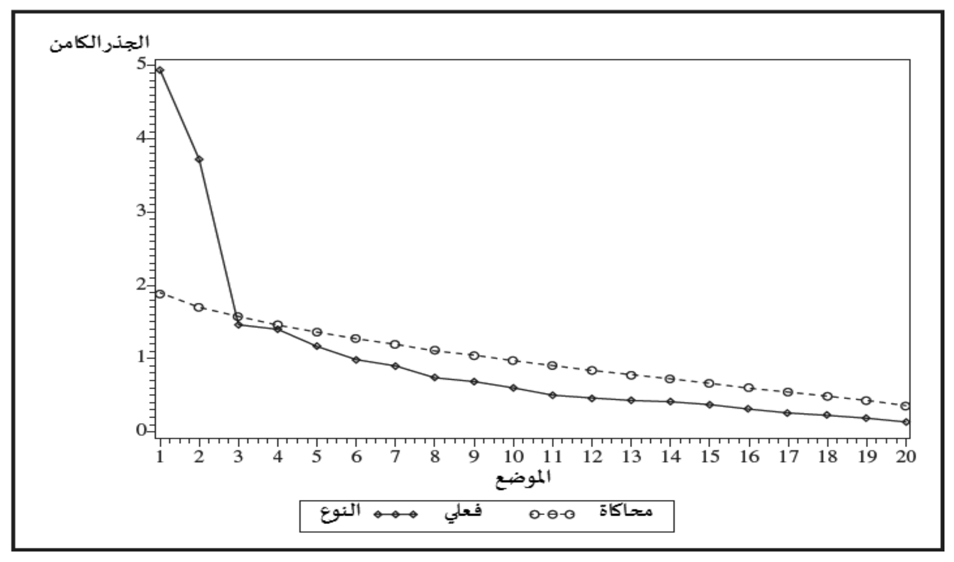
يوجد معيار إحصائي آخر، له شعبية متزايدة، يعتمد على التحليل الموازي parallel analysis (على سبيل المثال، Hayton, Allen, & Scarpello, 2004). المنطق خلف هذا النهج هو أن حجم الجذر الكامن للعامل الأخير المُحتفَظ به يجب أن يتجاوز الجذر الكامن الذي تم الحصول عليه من بيانات عشوائية في ظل ظروف مماثلة. بمعنى آخر، في دراسة حقيقية تتضمن التحليل العاملي، قل مثلاً، ٢٠ متغيراً مقيسًا على ٥٠٠ شخص، يجب أن يكون الجذر الكامن لأي عامل مُحتفَظ به أكبر من الجذر الكامن المقابل الذي تم الحصول عليه من بيانات تم إنشاؤها عشوائيًا مُرتَّبة لتمثيل ٢٠ متغيراً تم قياسها على ٥٠٠ شخص. تقوم برمجيات التحليل الموازي، والتي تتوفر على الإنترنت كوحدات ماكرو مطورة بواسطة المستخدم لمجموعة متنوعة من الحزم الإحصائية، بإنشاء عدد كبير من التوزيعات تتوافق مع حجم مجموعة البيانات الفعلية.



شكل (٦-٥): رسم الحصاة scree plot لا يتضمن مرفق elbow واضح

وبالتالي، في هذا المثال، سوف تولد الخوارزمية الكثير (يمكن للمستخدم عادةً تحديد العدد؛ ١٠٠٠غالباً العدد الافتراضي) من مجموعات البيانات العشوائية المكونة من ٢٠ متغيراً

و٥٠٠ شــخص. يتم استخراج أو اســتخلاص الجذور الكامنة لكل مجموعة من مجموعات البيانات المولدة، ويتم إنشــاء توزيع الجــذور الكامنة ضمن البرنامج للعامل الأول والعامل الثاني، وهكذا تســتمر العملية لبقية العوامل. لكل من هذه التوزيعات، يتم تحديد قيمة ممثلة (على سبيل المثال، الوسيط). يتم عرض المخرجات الرسومية، كما هو مبين في الشكل رقــم (٦-٦). يحتوي هذا الشــكل على خطــين يربطان النقاط المتعلقــة بالجذور الكامنة للعوامل المتتالية. الخط المســتقيم تقريباً المتدرج تنازلياً من اليسار إلى اليمين يمثل وسيط الجذور الكامنة (عبر مجموعات البيانات العديدة المولدة بالحاســوب) للبيانات العشوائية التي لا ترتبط المتغيرات فيها بالعوامل الأساســية الســائدة. الخط الآخر في الشكل هو رسم الحصــاة الفعلي، اســتنادًا إلى البيانات الحقيقية محل الاهتمام. ويشــار إلى عدد العوامل التي يجب الاحتفــاظ بها من خلال تقاطع خطي البيانات الفعلية والعشــوائية (المولدة). حجم الجذور الكامنة التي تقع فوق الخط الذي يمثل البيانات العشــوائية أكبر من المتوقع بالصدفة بمفردها (أي إنها ذات دلالة إحصائية). وتلك القيم التي تقع أسفل خط البيانات العشوائية ليست معنوية (على سبيل المثال، يمكن توقع حدوثهم بالصدفة البحتة).



شكل (٦-٦): شكل بياني من التحليل الموازي يوضح حجم الجذور الكامنة للعوامل المتتالية من البيانات المولدة (المحاكاة) والبيانات الفعلية

في المثال الموضح في الشــكل المصاحب، حقق العاملان الأولان من البيانات الفعلية قيمًا ذاتية أكبر من تلك للعاملين المقابلين اســتناداً إلى البيانات المولدة بشكل عشوائي. وبالتالي، فــإن هذين العاملين الأولين فقط يفيان بمعايير الاســتبقاء. لاحظ أن البيانات العشــوائية المولدة تســفر عن عدد من العوامل بجذور كامنــة أكبر من ١,٠، مما يدل على عدم كفاية قاعــدة الجــذور الكامنة. يُظهر فحص خــط البيانات الحقيقية أنه ســيتم أيضًا الاحتفاظ بعاملــين اســتنادًا إلى معيار اختبار الحصــاة الذي يبقي على العوامــل الواقعة فوق نقطة "المرفق" (التي تتوافق مع العامل الثالث) من الرسم البياني. ولأن وحدات الماكرو (برمجيات حاسوبية) لإجراء التحليل الموازي في مختلف الحزم الإحصائية (بما في ذلك SAS وSPSS) جاهــزة ومتاحة عبر الإنترنت دون قيود على اســتخدامها ودون مقابــل، فإن تطبيق هذه الطريقة كوســيلة لاختيار عدد العوامل التي يجب الاحتفاظ بها أصبح ممارســة شــائعة. ومــن خلال تجربتي، فإن التوجيه الذي تقدمه هذه الطريقة يتماشى مع الأحكام التي يتم إصدارها بناءً على معايير ذاتية مثل رسم الحصاة. ومع ذلك، فإن لها ميزة كونها ترتكز على معيار إحصائي، وبالتالي فهي تتمتع بقبول أوســع من الأســاليب الذاتية. وفقًا لذلك، أوصي باستخدام التحليل الموازي كأساس جوهري لتحديد عدد البنود التي يجب استخراجها.

تجدر الإشــارة إلى أنــه تم تطوير التحليل المــوازي لأول مرة للاســتخدام مع المكونات الرئيســية بدلاً من العوامل المشــتركة. وســتتم مقارنة هذه الأســاليب التحليلية المختلفة للعوامــل لاحقــاً. حاليــاً، نلاحظ فقــط أنه بينــما يتم حســاب المكونات الرئيســية من مصفوفــة الارتباط الأصلية، يتم حســاب العوامل المشــتركة من مصفوفــة ارتباط مخفضة reduced correlation matrix تحتــوي على تقديــرات تم ضبطها لإزالة تباين الخطأ من المتغيرات. تم توســيع التحليل الموازي لاســتخدامه مع العوامل المشتركة، ولكن كما لاحظ Timmermanو Lorenzo-Seva (2011)، فــإن عدد الجذور الكامنة للبيانات الحقيقية التــي تتجاوز تلــك الخاصة بالبيانات العشــوائية يميل إلى أن يكون أكــبر في حالة العوامل المشتركة منها في حالة المكونات الرئيسية. المنطق وراء التحليل الموازي تم تطويره مع وضع المكونات في الاعتبار. وبالتالي، في بعض الحالات، فإن عدد العوامل التي يسفر عنها التحليل الموازي التي ينبغي الاحتفاظ بها من العوامل المشــتركة قد يكون تقديرًا مبالغًا فيه للأبعاد الحقيقية لمجموعة البيانات.

قــد تظهر هناك مواقف عندمــا تدعم اعتباراتٌ جوهريةٌ (مثل قابلية تفســير العوامل الناتجــة) الاحتفاظ بالمزيد أو أقل من العوامل باســتخدام طريقة مثــل التحليل المتوازي. عندما يحدث هذا، نقترح تقديم نتائج الطريقة المستخدمة أولاً (على سبيل المثال، التحليل الموازي) مع الأساس المنطقي لاتخاذ القرار للابتعاد عن التوجيه الذي توفره هذه الطريقة. وإذا كانــت المســائل الجوهرية مقنعة، فقــد تكون هناك بعض الصعوبــة في إقناع القراء والمراجعين بأنه قد تم التوصل إلى قرار مناســب. من ناحية أخرى، إذا كانت الحجج التي تــم تنظيمها دعماً لتجاهل النتائج الأولية ليســت مقنعة، فقد لا يكون هناك ما يبرر رفض هذه النتائج.

تدوير العوامل:

الغرض من استخلاص العوامل هو بكل بساطة تحديد العدد المناسب من العوامل التي يجب فحصها. إن وضع المعلومات في التكوين أو الشــكل الأكثر فهماً ليس هو الهدف من استخلاص العوامل. العوامل الخام الغير مدارة unrotated هي تجريدات رياضية لا معنى لهــا. وكقياس تقريبــي، تخيل أنه قد طُلب منك وصف طول جميع الأشــخاص في الغرفة. وقررت أن تفعل ذلك عن طريق اختيار شخص بشكل اعتباطي، حيث تم اختيار جو بشكل عشــوائي، وقياس طوله، ووصف أي شــخص آخر بأنه أطول أو أقصر من الشخص المرجعي، جو، بعدة بوصات. لذلك؛ قد يكون أحد الأشخاص أطول من جو بثلاث بوصات وآخر أقصر منــه بمقدار بوصتين. في مثل هــذه الحالة، جميع المعلومات حــول الأطوال متوفرة بعدة طرق تقديمية للبيانات ولكن لم يتم تنظيمها بالطريقة الأكثر إفادة. ســيكون من الأســهل على الناس تفسير البيانات إذا تم تحويلها إلى نموذج يمكن فهمه بسهولة، مثل طول كل فرد في الغرفة، معبراً عنه بالأقدام والبوصات. يشبه دوران العوامل التحول من حيث إنه يقدم البيانات المتاحة مسبقاً بطريقة أسهل للفهم.

قبل محاولة تفســير العوامل - للتأكد من المفاهيم البنائية أو المتغيرات الكامنة المقابلة للعوامل، على أســاس البنود التي تم تحديدهــا مع كل عامل - عادةً ما يكون من الضروري إجراء تدوير للعوامل. يزيد دوران العوامل من قابلية التفسير من خلال تحديد مجموعات المتغــيرات التي يمكن وصفها في الغالب بمتغير كامن واحد (أي، البنود المتشــابهة من حيث

إنهـا جميعًا لها علاقـة قوية مع بعضها البعض، وبالتالي يتـم تحديدها إلى حدٍ كبير، بعامل واحد فقط). إن تدوير العوامل وقابلية التفسير الأفضل الناتجة عنه لا تتم بتغيير البنود أو العلاقات فيما بينها، ولكن عن طريق اختيار وجهات النظر التي يتم وصف البنود بناءً عليها.

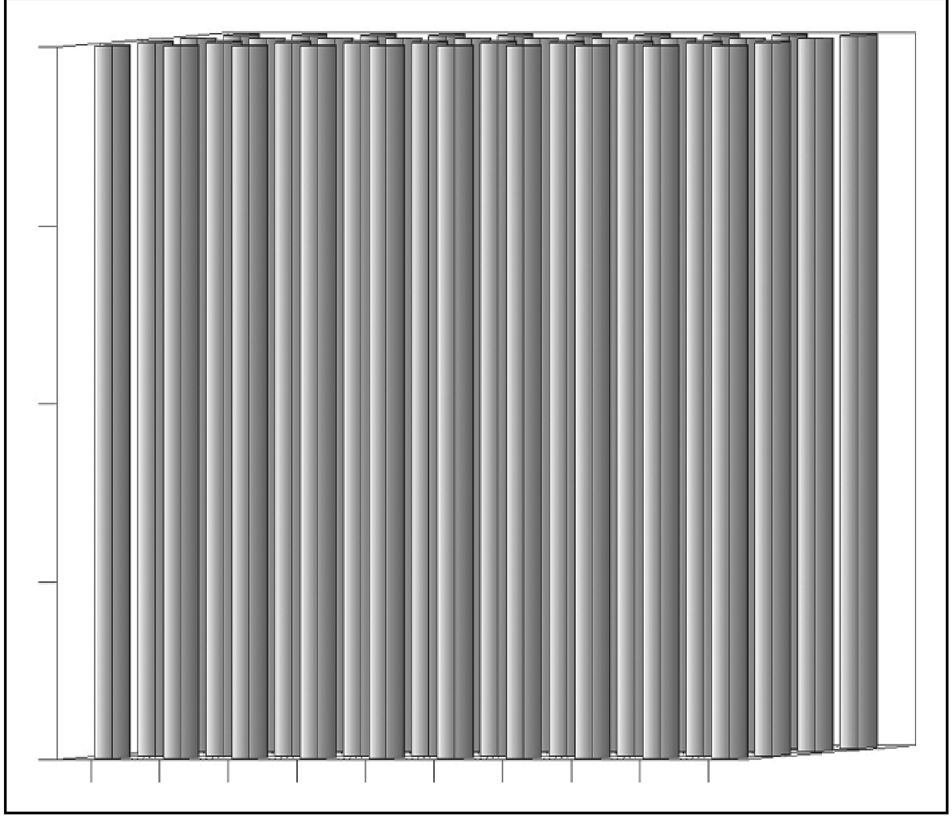
من الصعب فهم فكرة أن نفس مجموعة المعلومات يمكن أن تكون أكثر أو أقل وضوحاً بناءً على وجهة النظر التي يتم فحص تلك المجموعة من خلالها. في الأقسـام التالية، سيتم تقديم عدة أمثلة تهدف إلى توضيح هذه الفكرة.

تتشـابه أنماط العلاقات المتبادلة بين مجموعـة من البنود مع المواقع المادية في الفضاء. كلـما كان بندان أكـثر ارتباطاً، كانت العلامتـان الممثلتان لهذين البنديـن أقرب لبعضهما البعض. إذا قمنا بذلك للعديد من البنود، فإن المواقع الفعلية لعلاماتها سـتتخذ نمطًا يمثل أنماط الارتباط بين المتغيرات (من الأسـهل تصور ذلك إذا قصرنا أنفسـنا على بعدين). إن تخيل مواقع الأجسـام المادية التي تحددها القواعد الأساسـية هـو، بالتالي، طريقة أخرى للتفكير في البنود التي يتم تحديد ارتباطاتها بواسطة المتغيرات السببية الأساسية.

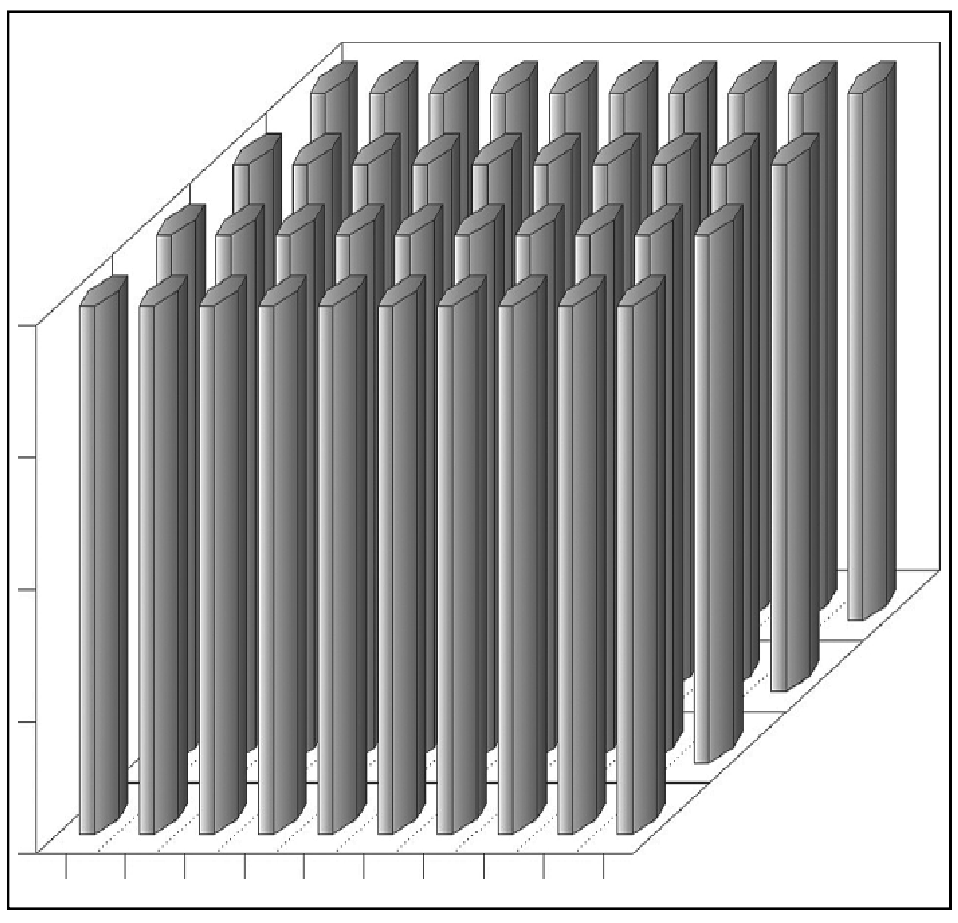
التشبيه الأول لتدوير العوامل:

كيـف يتيـح لنا الـدوران رؤية نمط بين المتغـيرات الذي كان دائماً هنـاك ولكن لم يكن واضحاً؟ كقياس، تخيل مجموعة من الأشياء المنظمة جيدًا، مثل مجموعة من الأعمدة مرتبة في سلسـلة من الصفوف المتوازية والمنظمة. من الممكـن الوقوف في مواقع معينة واختيار زوايا رؤية معينة بحيث يكون ترتيب الأشياء في الصفوف والأعمدة مخفياً بشكل كامل. ومع ذلك، تغيير زاوية الرؤية يمكن أن تكشـف عن الترتيب الأساسي. فكر في كيفية ظهور ترتيب أربعـة صفوف من ١٠ ركائز pillars عند عرضها من وجهات نظر مختلفة. إذا لم يكشـف خـط الرؤية عن أيٍ مـن المحاور الطبيعيـة للترتيبات الخطية، فقد تظهر الأعمدة بشـكل عشـوائي. المشـاهدة من موقع أعلى، أو التحرك بضعة أقدام يمينًا أو يسارًا، أو مجرد تحوّل نظر المرء أفقيًا إلى خط الرؤية يمكن أن يؤدي إلى الكشف عن انتظام الركائز وإظهار ترتيبها، ويوضح الشكلان رقم (٦-٧) ورقم (٦-٨) هذه النقطة. هذه الأشكال تم إنشاؤها عن طريق رسم مواقع للأعمدة، وحرفياً، تغيير موقع الرؤية الذي من خلاله تم عرض الأعمدة. وبالتالي، فإن كلا الشكلين يمثلان نفس تكوين الأعمدة ويختلفان فقط في كيفية مشاهدتها. في الشكل

الأول من هذه الأشكال، من الصعب في أحسن الأحوال تمييز أي تنظيم لترتيب الركائز. يقع خط الرؤية الذي يمر خلال المسافات بين الركائز على ركائز أخرى أبعد، جاعلاً التجمع يظهر ككتلة غير منظمة من الأشــياء. يمثل الشــكل الثاني من هذه الأشكال وجهة نظر بديلة. كل ركيزة، كما يوضح الشكل (٦-٨)، تشارك صفًا (وعمودًا) مع ركائز أخرى. وبالتالي، فإن جميع الركائز في صف معين لها شيء مشــترك - فهي تمتلك ســمة مشتركة (عضوية في نفس الصف) لم تكن واضحة من موقع الرؤية الســابق. إن مجرد تغيير زاوية الرؤية جعل شيئًا ما يتعلق بطبيعة الأشياء واضحاً كان في منظورنا الأولي غامضاً. دوران العوامل يشبه هذه المحاكاة في أنه يحاول توفــير "وجهة النظر" التي من خلالها يصبح الهيكل التنظيمي للبيانات - الطرق التي تشترك بها البنود في بعض الخصائص - واضحاً.



شكل (٦-٧): ترتيب منظم للركائز التي يتم رؤيتها من منظور يحجب ذلك الترتيب



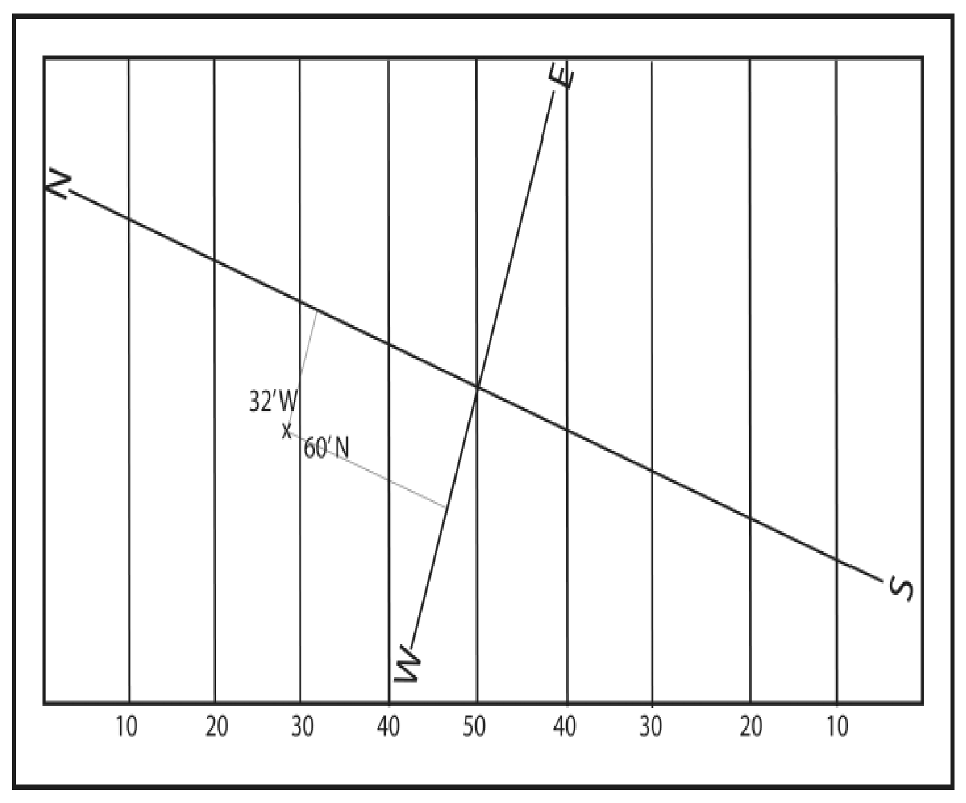
شكل (٦-٨): ترتيب منظم للركائز التي تتم رؤيتها من منظور يكشف عن ذلك الترتيب

التشبيه الثاني لتدوير العوامل:

تجدر الإشــارة إلى أنه، مع العدد الصحيح من الخطوط العمودية المرجعية، يمكن للمرء تحديــد موقع الأشــياء بغض النظر عن كيفية توجيه هذه الخطــوط المرجعية. مثال ثنائي الأبعاد مثل ملعب كرة قدم كبير فارغ يمكن اســتخدامه للتوضيح. لجعل هذا المشهد أكثر إثــارة للاهتمام قليــلاً، دعونا نفترض أن المدرب يســتخدم مباراة فيديــو، تم تصويرها من كامــيرا علوية، لشرح اللعب لأعضــاء الفريق. علاوة على ذلك، افــترض أن المدرب قد قام

بتجميد الصورة في لحظة مثيرة للاهتمام بشكل خاص أثناء اللعب لتوضيح نقطة معينة وأن هذه الصورة المجمدة يتم عرضها على شاشـة كمبيوتر كبيرة تتيح للمدرب الرسم فوق تلك الصورة. يمكن للمدرب رسـم خط مسـتقيم من خلال صورة ملعب كرة القدم في أي اتجاه ثم وضع خط ثانوي عمودي على الأول. على سبيل المثال، قد يختار المدرب خطًا واحدًا يمتد تماماً شـمالًا وجنوباً وآخر يمتد على وجه التحديد شرقاً وغرباً، والتي قد تسـمح له بوصف أي موقع في الحقل باستخدام إحداثيات تحديد المواقع الشاملة استنادًا إلى خط العرض من الشـمال إلى الجنوب وخط الطول من الشرق إلى الغرب. باسـتخدام هذين الخطين، يمكن للمـدرب تحديد موقع اللاعب الـذي يمتلك الكرة: «باسـتخدام إحداثيات تحديد المواقع الشـاملة لوسـط الملعب كمرجع، يكون حامل الكرة عند تقاطـع نقطة محددة من خلال التقدم شمالًا من وسط الملعب لمسافة ٦٠ قدمًا ثم المضي قدمًا غربًا ٣٢ قدمًا إضافيًا». هذا سَـيُحدد لاعب كرة القدم في مكان معين. يمكن للمدرب تحديد الموقع نفسـه باسـتخدام تعليمات معدلة بشكل مناسب بناءً على أي خطين عموديين مرسومين في ملعب كرة القدم. لذلـك، فإن اتجاه الخطـوط هو أمر اعتباطي فيما يتعلق بكفايـة وصف موقع معين. أي مجموعـة من خطين عموديين لهما نفس الفائدة المعلوماتية في تحديد مكان معين مثل أي مجموعة أخرى.

بدلاً من ذلك، بدلاً من استخدام المحورين شمال - جنوب وشرق - غرب، يمكن للمدرب اسـتخدام الخطوط الجانبية وخطوط المرمى للملعب كمحاور متعامدة أساسـية. مثل أي مجموعة أخرى من الخطوط المتعامدة، يمكن استخدام هذه المحاور كمراجع لتحديد موقع أي نقطـة في الملعـب. ومع ذلـك، فهي تتمتع بميزة الاسـتفادة من ميـزة ذات مغزى في الملعب — تنظيمها كشبكة محددة بخطوط متعامدة (أي خطوط ساحة الملعب والخطوط الجانبيـة). وبالتالي، فإن إخبار شـخص ما بأن حامل الكـرة كان ٤ أقدام داخل خط الـ ٣٠ ياردة وحوالي ٦٨ قدمًا من الخط الجانبي لملعب الفريق المضيف، قد تكون وسـيلة أسـهل لتحديد موقع هذا اللاعب من اسـتخدام الخطوط المرجعية الممتدة بين الشمال والجنوب والـشرق والغـرب. النقطة المهمة هـي أنه في حين أن أي مجموعة مـن المحاور المرجعية يمكنها تحديد موقع حامل الكرة بشـكل لا لبس فيه، فإن المجموعة القائمة على الخصائص الطبيعية لملعب كرة القدم تبدو أكثر ملاءمةً وشموليةً (انظر الشكل رقم ٦-٩).



شكل (٦-٩): وصف لملعب كرة قدم معلم بخطوط الميدان، والخطوط الجانبية، وإحداثيات بين الشمال - الجنوب والشرق - الغرب

الآن وقد رأينا كيف يمكن استخدام المتجهات أو الخطوط المرجعية لتحديد موقع مكاني، يجب أن نفحص كيفية ارتباط البنود بخطوط مرجعية مماثلة. ومع ذلك، قد يكون من الصعب التبديل بين التفكير في البنود ومحتواها من ناحية والتفكير في المتجهات المجردة التي تحدد المواقع من ناحية أخرى. في كثير من الأحيان، الأمثلة التي تربط البنود الأكثر مألوفية بتمثيلها المكاني على طول المتجهات يجب التغاضي عنها، لأن الأشياء التي نقيسها كعلماء اجتماع وسلوك لا تتناسب عادةً مع ترجمة مباشرة بالحدود المكانية. لسد هذه الفجوة، سأقدم مثالًا مختلقاً (غير حقيقي) استنادًا إلى بنود يمكننا ربطها بالمواقع المادية.

التشبيه الثالث لتدوير العوامل:

امتداداً لمثال كرة القدم لدينا، سـوف نفترض دراسـة تسـتخدم بنود الاستبانة كوسيلة لوصـف موقع لاعب كرة القدم (الشـخص الـذي يحمل الكرة) في الملعب. وبشـكل أكثر تحديداً، دعنا نفترض أنه يُطلب من مجموعة من المُسـتطلَعين مشـاهدة فيديو عن لعبة كرة القدم، وعند نقطة ما، يتم تجميد الصورة ويطلب من المجيبين الإجابة على سلسلة من ثمانية بنود باستخدام مقياس ليكرت المتدرج والذي يتراوح من ١ = لا أوافق بشدة إلى ٦ = أوافق بشدة. هذه البنود كالتالي:

A- حامل الكرة في موضع تهديف سهل.

B- قد يتم إجبار حامل الكرة على الخروج من الحدود سريعاً.

C- قد يُجبر حامل الكرة إلى منطقة النهاية الخاصة به.

D- حامل الكرة لديه مساحة كبيرة لأي من الجانبين.

E- إذا تم إيقاف حاملة الكرة الآن، شخص ما على الخطوط الجانبية قد يكون في خطر.

F- أمام حامل الكرة طريق طويل قبل عبور خط المرمى.

G- حامل الكرة داخل النطاق الميداني للمرمى.

H- يجب أن يكون حامل الكرة حريصاً على عدم تجاوز الخط الجانبي.

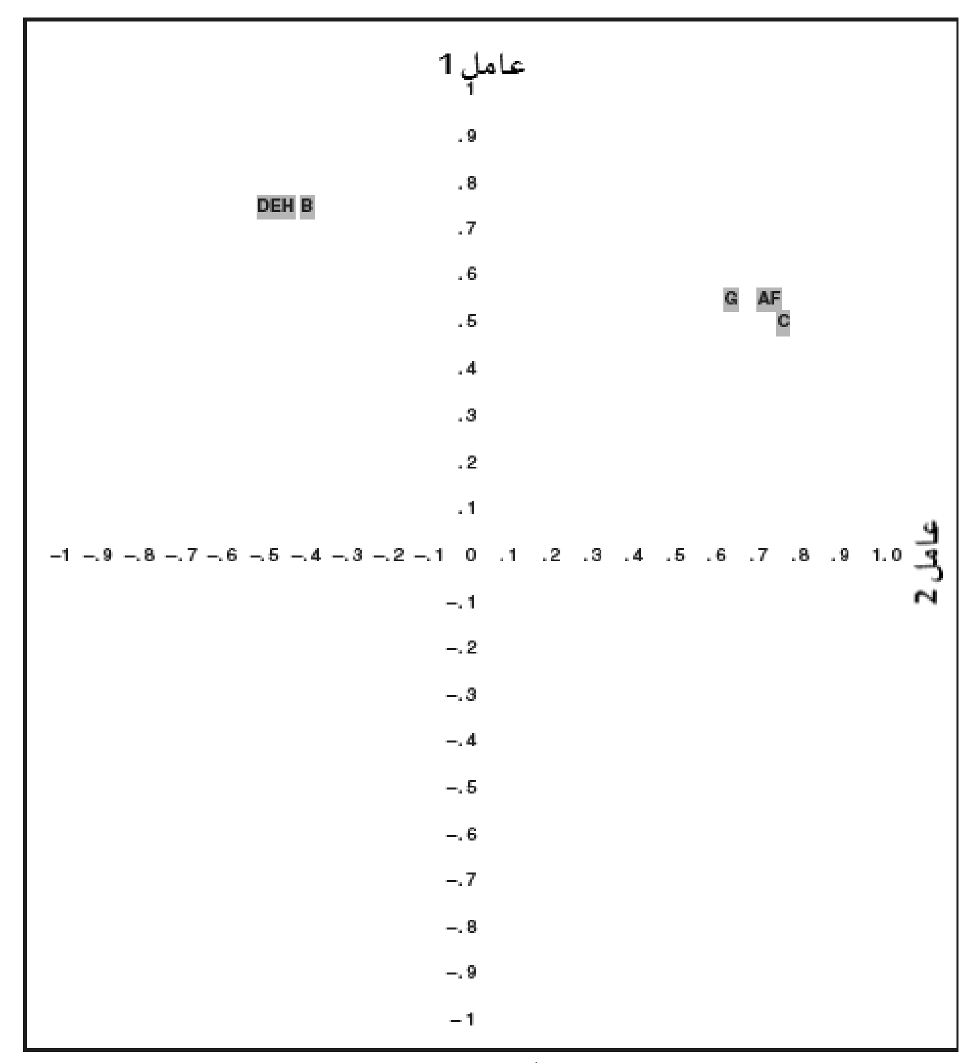
قبل المتابعة مع هذا المثال، دعونا نتوقف للحظة بشكل أكثر رسمية للتعرف على كيفية دوران العوامل. عندما يكون لدينا عاملان، فإنه يمكن تحقيق الدوران بيانياً. ويمكن للمرء الحصول على الشـكل الانتشـاري **scatterplot** الذي يمثل قوة الارتباطـات بين أزواج من البنـود عن طريق القـرب من هذه البنود وتشـبعات **loadings** تلك البنود على متجهات مرجعية اعتباطية مقابلة لاثنين من العوامل الغير مدارة. يمكن للمرء إجراء دوران رسومي من خلال تدوير تلك المحاور حتى تسقط على مجموعات البنود. بعد ذلك، يتم استخدام المحاور المرجعية لإنشـاء إحداثيات تحدد الموضع الدقيق لكل بند ممثَّل من حيث موقعه بالنسـبة إلى الأطوال على امتداد المحورين. ويمكن بعـد ذلك تحويل هذه الإحداثيات إلى تشبعات لكل بند على هذين العاملين. ولكن طرق الدوران غير الرسومية الآلية القائمة على

المعايير الرياضية تســتخدم على نطاق أوسع. على سبيل المثال، تسعى أساليب مثل التباين الأعظم Varimax إلى زيادة تباين التشــبعات المربعة لكل بند. التشــبعات هي عبارة عن ارتباطــات بين كل بنــد وكل عامل بناءً على اتجاه المتجهــات (أي الخطوط المرجعية) التي تحدد العوامل. وبناءً على الكيفية التي يتم بها توجيه المتجهات، ستتباين التشبعات، تمامًا كــما تتباين القيــم التي تحدد موقع لاعب كرة قدم في الملعــب أو الإحداثيات التي تصف موضع نقطة ما على الشــكل الانتشاري اعتمادًا على المحاور المرجعية المستخدمة. سيكون هذا التباين أكبر عندما تكون بعض التشــبعات المربعة كبيرة نســبيًا والبعض الآخر صغيراً نسبيًا. وبالتالي، فإن تلبية معيار تعظيم تباين التشبعات المربعة له تأثير في إنتاج مجموعة متفاوتة القيم القصوى من التشــبعات لكل بند. العثور على اتجاه للمحاور المرجعية التي ستمنحنا تشبعات متفاوتة القيم القصوى والقيم الأكبر لتباينات التشبعات المربعة سيؤدي إلى إيجاد الدوران الذي من المحتمل أن يكون مفيداً. ما نسعى إليه، بالطبع، هو نوع معين من التفاوت بحيث يكون تشبعاً واحداً (أي، تشبع البند على العامل الأساسي) كبيراً في حين أن جميع التشبعات الأخرى لهذا البند (على سبيل المثال، التشبعات الخاصة به على العوامل الثانوية) صغيرة (من الناحية المثالية، قريبة من الصفر). عندما يتحقق ذلك، يمكننا وصف البند بأنه يتأثر بشــكل أساسي بالعامل الفردي الذي يتشبع عليه بشكل كبير - وهي الحالة التي تدعى الهيكل البسيط. إن وصف بندٍ من خلال عامله الأساسي فقط يقوم بعمل جيد لالتقاط الطبيعة الأساسية للمعلومات التي يحملها البند. هذا هو هدف الدوران.

الآن، دعونــا نعــود إلى البنود التي تصف موقع حامل الكــرة في الصورة المجمدة للعبة كرة القدم. لقد تم تحليل هذه البنود الثمانية (بناءً على بيانات مزيفة) لأغراض توضيحية. وكانــت الجــذور الكامنة الثلاثة الأولى ٣,٦٧ و٢,٧٨ و٠,٥٩. بنــاءً على الحقيقة أن القيمتين الأوليين أكبر بكثير من ١,٠ بينما الثالثة أصغر بكثير من ١,٠، وتم اســتخراج عاملين. حيث طلــب من برنامــج SAS إجراء دوران بأســلوب التباين الأعظــم Varimax على العوامل. الشــكلان رقمي (٦-١٠) و(٦-١١) يوضحان رســومات العوامل باستخدام برنامج SAS قبل وبعد الدوران، على التوالي. وهذه تحدد موقع البنود فيما يتعلق بالعاملين بناءً على تشبع كل بند على هذين العاملين.

يقوم SAS بطباعة رسـومات بيانية خاصة للعوامل بخط يسـتخدم العرض المتسـاوي لكل حرف ولا يسـتطيع التقدير interpolate بين خطوط من النصوص. وبالتالي، لا يمكن تحديد موقع النقاط بالنسبة لبعضها البعض أو إلى متجهات العوامل بدقة تامة. ومع ذلك، حتـى مـع هذا القيد، فإن الرسـومات البيانية إيحائية تماماً. في الشـكل رقم (٦-١٠)، يتم تجميع البنـود (ممثلة بأحرف داكنة في مربعات مظللة لرؤية أكثر وضوحًا) في مجموعتين، إحداهـما تشـتمل على البنود B وD وE وH والأخرى، البنـود A وF وC وG. القرب بين البنود يشـير إلى قوة الارتباط في هذا الشـكل. لذلك، أي بندين يقعان بالقرب من بعضهما البعـض مرتبطان بدرجة كبيرة أكثر من أي زوج بعيد. وبالتالي، يشـير وجود مجموعتين في هذا الشكل إلى أن البنود الموجودة في كل مجموعة ترتبط ارتباطًا وثيقًا ببعضها البعض وأن البنـود الموجودة في مجموعات مختلفة غير مرتبطة إلى حـدٍ كبير ببعضها البعض. المواقع بالنسـبة إلى المحاور هي أيضاً مفيدة وذات صلة خاصة بمسـألة الـدوران. تقع مجموعة البنود A-F-C-G على بعد متسـاوٍ تقريبًا من المتجهين الممثلين للعوامل. أي إن تشـبعاتها على كل عامل تقريبًا بنفس الحجم. مجموعة البنود الأخرى، على الرغم من أنها أقرب إلى حـد ما مـن المتجه الرأسي منها إلى المتجه الأفقي، فإنها لا تزال على مسـافة من أي منهما. تحتوي البنود الموجودة في المجموعة الثانية على تشـبعات أعلى على العامل ١، ولكن أيضًا تشبعات (سالبة) كبيرة إلى حد ما على العامل ٢. وبالتالي، سيكون من الصعب القول إن أي من مجموعتي البنود يعكس بشكل أساسي مستوى واحداً فقط من العوامل.

في المقابل، في الشـكل الذي تم إنشـاؤه بعد أن تم تدويـر العوامل، نلاحظ أن كلاً من المجموعتـين قريبة جدًا من متجه واحد فقط. المجموعـة الأولى من البنود (B-D-E-H) لديها الآن تشـبعات عاليـة (تقريبًا من ٠,٨ إلى ٠,٩) على العامـل ١ (يمثله المحور الرأسي) وتشـبعات منخفضة على العامل ٢ (أقل من ٠,٢). المجموعـة الثانية (A-F-C-G) قريبة مـن العامل ٢ (المحور الأفقـي)، وكل البنود مجتمعة تمثل قيم عالية على العامل ٢ (تقريبًا في حدود ٠,٨٠ إلى ٠,٩٠) وقيم منخفضة على العامل ١ (حوالي ٠,١٠).

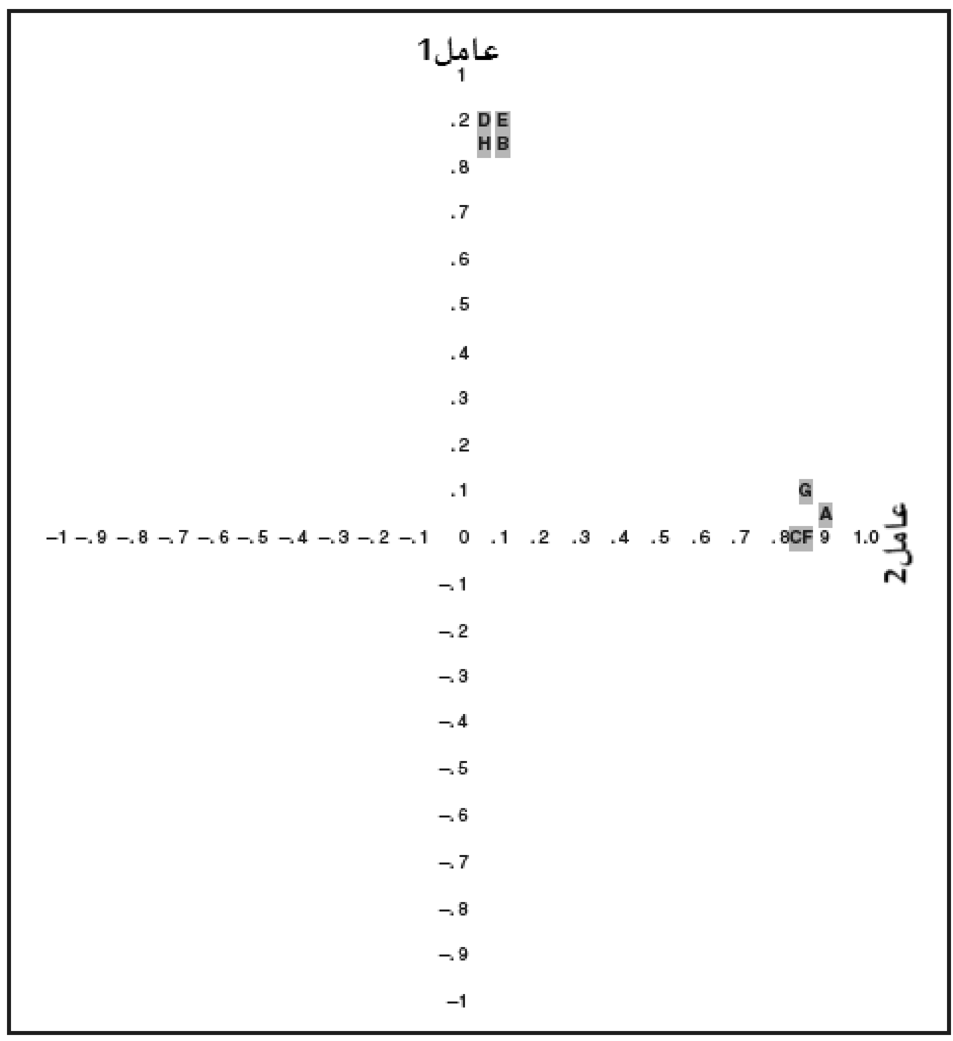


شكل (٦-١١): العناصر الممثلة بيانياً بواسطة التشبعات على عاملين قبل الدوران

لذلــك، بعد الــدوران، لدينا تقريب أفضل بكثير لهيكل بســيط مــن توزيع البنود على العوامل - أي، كل بند يتشبع بشكل أساسي على عامل واحد.

إذا نظرت بعناية إلى المواقع النســبية لمجموعتي البنود في الشكلين المختلفين، قد تكون قــادرًا عــلى التمييز على أنهما متشــابهان إلى حد كبير (في حدود الرســم البيــاني لـ SAS).

مجموعــة البنــود A-F-C-G تقع حوالي ٩٠ درجة في اتجاه عقارب الســاعة من المجموعة B-D-E-H في كلا الشــكلين. هــذا كما يجب أن يكون، لأن الــدوران لا يغير العلاقات بين المتغيرات. لو كانت البنود مرســومة على ورقة شــفافة والعوامل على ورقة شــفافة أخرى، فإنــه يمكنك تدوير الورقة المحتوية على البنود من الموضع الموضح في الشــكل رقم (٦-١٠) إلى الموضع الموضح في الشكل رقم (٦-١١) الخاصة بالعوامل.



شكل (٦-١١): العناصر الممثلة بيانياً بواسطة التشبعات على عاملين بعد الدوران

(حقيقة أن المحاور الأفقية والعمودية لها مقاييس مختلفة قليلاً بسـبب عدم الدقة في كيفية عرضSAS للرسـوم البيانية، قد تجعل مسافات المجموعات من الأصل أقل تشابهًا في الشكلين مما هي عليه بالفعل عدديًا). في الحقيقة، إن الدوران يعيد فعلاً توجيه العوامل، ولكـن بدلاً من عرض محاور العامل بزوايا غريبة، تقوم رسـومات ما بعد الدوران بشـكل أساسي بتدوير مواقع البنود في اتجاه عقارب الساعة بدلاً من تدوير المحاور في اتجاه عكس عقارب الساعة.

بعـد تدوير العوامل، أصبح من الممكن الآن تفسـير كل منها مـن خلال النظر في ماذا تشترك البنود التي حققت تشبعات عالية على كل من العوامل. البنود التي تشكل العامل الأول جميعها تتعلق بالقرب من الخطوط الجانبية للملعب، وتلك التي تشكل العامل الثاني تتعلق بالقرب من خطوط المرمى. وبالتالي، قد نفسر العامل ١ على أنه القرب من الخطوط الجانبية والعامل ٢ على أنه القرب من خطوط المرمى. لقد سمح لنا الدوران بتجميع البنود بطريقة سهلت التعرف على السمة العامة المشتركة بين البنود داخل كل مجموعة.

باختصـار، إن أهـم النقاط التي تدل عليها هذه الأمثلة حـول دوران العوامل هي كما يـلي: (أ) معرفـة العدد الصحيح للعوامل (عندما يكون هنـاك أكثر من عامل واحد)، رغم أنه ضروري، لا يكشـف عن طبيعة تلك العوامل. عملية اسـتخراج العوامل الأولية تتناول فقط عدد العوامل المطلوبة، وليس طبيعة تلك العوامل بأي معنى جوهري. (ب) في حالة غيـاب دوران العوامل المتعددة، فـإن أي نمط ذي معنى وقابل للتفسـير بين البنود يمكن حجبه. الحل الأوَّلي لاستخلاص العوامل يتجاهل محتوى البنود، وعادةً ما تكشف تجمعات البنود بناءً على عوامل غير مدارة القليل عن تلك البنود. ويوضح تصوير الركائز من منظور غـير ملائم ومن ثم منظور ملائـم هذه النقطة. (ج) يمكن لأي حل تحليلي يشـتمل على العدد الصحيح من العوامل تحديد العلاقات بين البنود (أو، إذا وصفناها مكانيًا، أي مواقع البنود بالنسـبة لبعضها البعـض) بالإضافة إلى غيرها. تعمل المحاور الشـمالية - الجنوبية والشرقيـة - الغربية تمامًا كما تعمل محاور المرمى - المرمى والخطوط الجانبية - الخطوط الجانبية لتحديد أي موقع في الملعب. (د) الغرض من تدوير العوامل هو إيجاد اتجاه معين للمحاور المرجعية التي تساعدنا على فهم البنود بعبارات بسيطة. يتم تحقيق ذلك عندما

يمكن شرح معظم محتويات البنود باســتخدام أقل عدد ممكن من المتغيرات - من الناحية المثالية، واحد فقط. (هـ) يمكن أن يســهل نمط العوامل المدارة تحديد ما هو مشــترك بين البنود داخل عامل ما وبالتالي اســتنتاج ما هو العامل الســببي الأســاسي الذي يحدد كيفية الإجابة عــلى البنود. كانت المجموعتان من البنود المتعلقة بموقع حامل الكرة في لعبة كرة القدم مختلفة بشــكل واضح. مجموعة واحدة تتعلق فقط بالقرب من الخطوط الجانبية، في حــين أن المجموعة الأخرى تتعلق فقط بالقرب من خطوط المرمى. يمكننا أن نســتنتج، عــلى الأقل مؤقتًا، أن تصورات القرب من الحــدود المختلفة لملعب كرة القدم تحدد كيفية تقييم المجيبين لمجموعتي البنود. بالطبع، ستكون هناك حاجة إلى معلومات تحقُّق إضافية لدعم هذا الاستدلال.

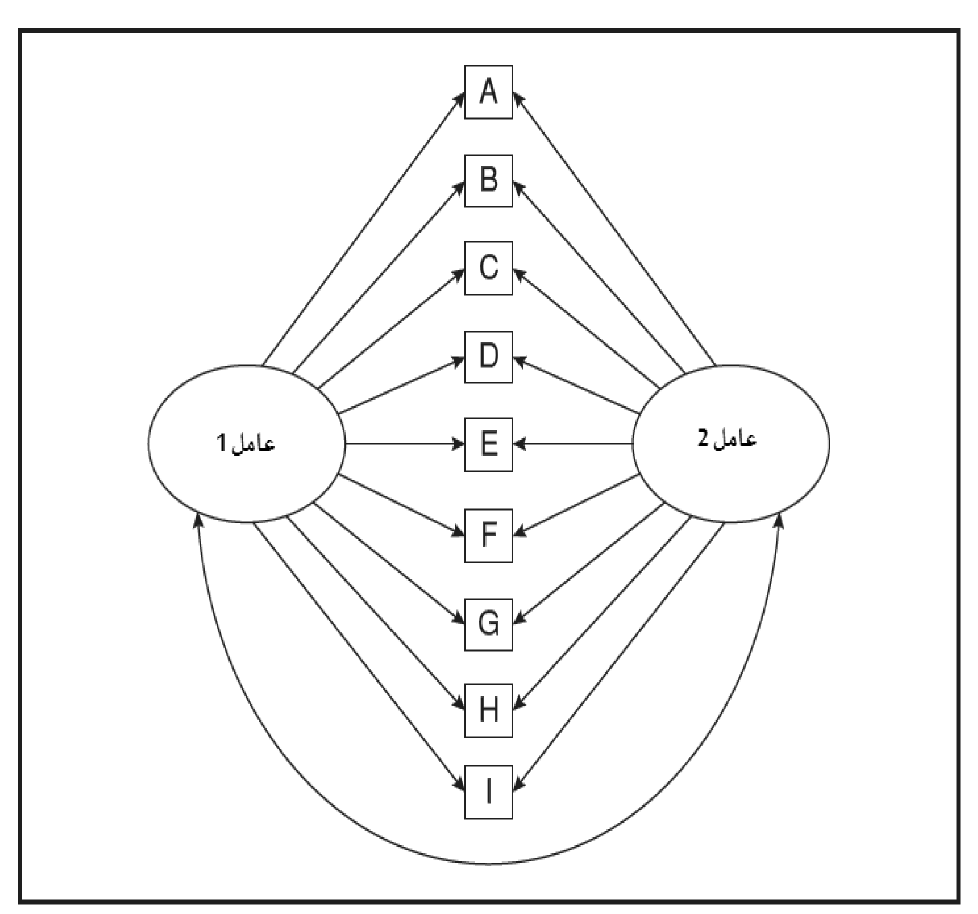
التدوير المتعامد مقابل التدوير المائل:

تســتند جميع الأمثلة حتى الآن إلى خطوط مرجعية متعامدة مع بعضها البعض. هذا يتوافق مع العوامل التي هي مســتقلة إحصائيًّا عن بعضها البعض (أي غير مترابطة). مثل هــذه العوامل توصف بأنها متعامدة orthogonal. لا يوفر الموقع على طول خط (محور) مــا أي معلومات تتعلق بالمعلومات الموجــودة على طول الخط الآخر عندما يكون الخطان متعامدَين. على سبيل المثال، معرفة مدى بعد أي شخص شمالاً لا تعطي مؤشرًا على معرفة مدى بعده غرباً، لأن هذين الاتجاهين متعامدان مع بعضهما البعض. وبالمثل، معرفة مدى بُعــد اللاعب عن خط المرمى لا تعطي مؤشراً على قربه من الخط الجانبي. ومع ذلك، من الممكن الســماح للعوامل بالارتباط، وبالتالي، فإن المحاور التي تمثلها بشكل بياني تكون غير متعامدة. لنفترض، على سبيل المثال، أننا اخترنا تحديد مواقع في ملعب كرة القدم كما هو موضح في الشــكل (٦-٩) باستخدام خط يمتد من منطقة النهاية إلى منطقة النهاية كمحور وخــط يمتــد شرقًا وغربًا مثل الآخر.التحــرك في اتجاه غربي يعني أيضًــا التحرك نحو إحدى مناطق النهاية. أي أنه لا يمكن للمرء أن يتحرك مباشرةً شرقًا أو غربًا دون تغيير مسافته عن خطوط المرمى. البعدان يرتبطان إلى حدٍ ما.

بالمثل، يمكن أن يسمح دوران العوامل للمحاور المرجعية (والعوامل التي تتوافق معها) لتكــون مرتبطة، وبالتالي، لا تكون متعامدة مكانيًّا. يشــار إلى هــذا الدوران على أنه مائل

oblique بـدلاً مـن المتعامد. قد يكون الـدوران المائل مفيدًا عندمـا يُعتقد أن المتغيرات الكامنـة ترتبط ببعضها البعض. يتطلب هدف الهيكل البسـيط البنود التي يمكن تصنيفها بشـكل مفيد فيما يتعلق بفئة واحدة فقط. أي، يجب أن يكون كل بند "عن" شيء واحد فقـط، وبالتالي، يتشـبع load على عامل واحد فقط. إذا كانـت المتغيرات مرتبطة إلى حدٍ ما ولكن العوامل التي تمثلها أُجبرت إلى أن تكون مسـتقلة تمامًا بسبب القيود التي تفرضها طريقة تحليل العوامل، فقد يكون من المستحيل تحقيق هذا الهدف. أي إن أكثر من عامل قد يرتبط ببعض أو كل البنود بسـبب الارتباط بين العوامل؛ نحن محدودون في قدرتنا على التقريب للهيكل البسيط (انظر الشكل رقم ٦-١٢).

بالعـودة إلى مثال خصائص زميل العمل السـابق، إذا كان "الاجتهـاد" و"الاعتمادية" مترابطـين حقًا، فمن المحتمل أن يتقاسـم بند عن أحدهما بعـض التباين مع الآخر أيضًا. الأشخاص الأكثر اجتهاداً سيكونون أيضًا أكثر اعتمادًا إلى حدٍ ما (بدرجة أكبر أو أقل استنادًا إلى قـوة الارتباط بين السـمتين). إن فرض حل متعامد عـلى البيانات التي تعكس هذين المتغيرين سـيجعل من الصعب للغاية على البند أن يتشـبع بشكل قوي على متغير واحد وبشـكل ضعيـف على الآخر. ومع ذلك، إذا سـمح للعاملين بأن يكونـا مترابطين إلى حدٍ ما، تكون الحالة مشـابهة تقريبًا للاسـتنتاج التالي: "الاجتهاد" و"الاعتمادية" يُفهم أنهما يرتبطان ببعضهما البعض. وقد تم التعامل مع هذه الحقيقة من خلال السـماح للعوامل بالارتباط. الآن، إلى جانب ذلك، إلى أي من هذه العوامل يرتبط البند المعني بشدة؟ إذا لم يتم إجبار المحاور التي تمثل عامل "الاجتهاد" وعامل "الاعتمادية" على أن تكون متعامدة، فمـع زيادة الدرجات على أحد المتغيرات، يمكن اسـتيعاب درجة ما من الزيادة المصاحبة في المتغير الآخر.

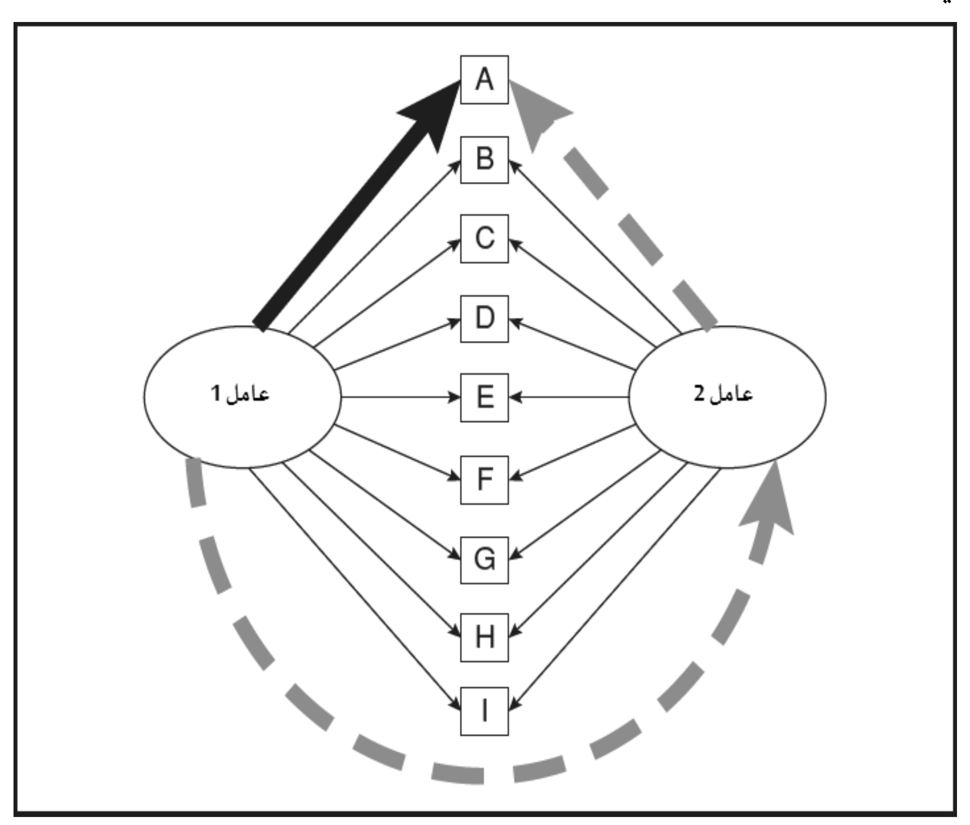


شكل (٦-١٢): نموذج بعاملين مع السماح للعاملين بالارتباط

إذا كان الواقــع يفــرض أن البند الذي يتأثر بشــكل كبــير بـ «الاجتهاد» يتأثــر أيضًا بـ «الاعتماديــة»، فإن الســماح للعوامل التي تمثل تلــك المفاهيم البنائيــة بالترابط يمكن أن يســتوعب تلك الحقيقة بشــكل أكــثر دقة. قد يكون البنــد الذي يرتبط ارتباطًــا وثيقًا بـ "الاجتهاد" ولكنه يرتبط بشــكل معتدل بـ "الاعتمادية" قادرًا على التشــبع بقوة على أول تلك العوامل وبشكل ضعيف على الثاني إذا كانت العوامل غير مقيدة بالاستقلال الإحصائي. بمعنى، حقيقة أن هذا البند له تأثير سببي ناشئ عن "الاعتمادية"، يتم استيعابه من خلال السماح للعاملين بالترابط مع بعضهما البعض. يوفر هذا مسارًا غير مباشر من "الاعتمادية"

إلى البند من خلال "الاجتهاد" (انظر الشكل رقم ٦-١٣)، مما يزيل الحاجة إلى ارتباط البند مباشرةً بـ "الاعتمادية" وبالتالي التشبع المتقاطع على كلا العاملين.

ما يتم فقدانه عندما يتم تدوير العوامل بشكل مائل (غير متعامد) هو أناقة وبساطة الأبعاد غير المترابطة. إحدى المزايا الجذابة للعوامل غير المترابطة هي أن تأثير العوامل مجتمعة عبارة عن حاصل المجموع البسيط لتأثيراتها منفصلة. مقدار المعلومات في قيمة بند معين الذي يفسره أحد العوامل يمكن إضافته إلى مقدار المعلومات الذي يفسره عامل آخر للحصول على إجمالي مقدار المعلومات المفسرة بالعاملين مع بعضهما البعض. في العوامل المائلة، الحال ليس كذلك.



شكل (٦-١٣): نظرًا للارتباطات بين العوامل، يؤثر العامل ١ على البند A بشكل مباشر (المسار الصلب الداكن) وغير مباشر (المسار المتقطع الفاتح)

ولأن العاملين مترابطان، يوجد تكرار في المعلومات الواردة في هذين العاملين. بالنسبة إلى بند يتعلق بكل من العاملين «الاجتهاد» و«الاعتمادية»، فإن مقدار التباين الذي يفسره هـذان العاملان معًا أقل من مجموع تباين العوامل منفصلة. بعض المعلومات، وربما قدرٌ كبيرٌ منها، التي يستأثر بها أحد العوامل تتداخل مع المعلومات التي يستأثر بها العامل الآخر. محصلة المجموع البسيط للمعلومات ستتضمن المعلومات المتداخلة مرتين، والتي لن تعكس بدقة التأثير الكلي للعوامل على ذلك البند.

ومن المضاعفات الأخرى للعوامل المترابطة التعقيد المضاف للعلاقات السببية بين البنود والعوامل. عندما تكون العوامل مستقلة، تكون العلاقة الوحيدة بين عامل وبند معين مباشرة. ستؤدي التغييرات في مستوى العامل إلى تغييرات في البند على امتداد مسار سببي مباشر واحد. غير أن هذا ليس هو الحال عندما تكون العوامل مترابطة. إذا كان (كما هو موضح سابقًا عند مناقشة بند متأثر بـ "الاجتهاد" و"الاعتمادية") يؤثر كلاً من عاملين افتراضيين على البند A، على سبيل المثال، وكان العاملان مترابطين، فإن العاملين يمارسان تأثيراً غير مباشر وكذلك مباشر على البند A. بمعنى، يمكن للعامل ١ التأثير على العامل ٢، ومن خلال العامل ٢، التأثير بشكل غير مباشر على البند A. هذا بالإضافة إلى التأثير المباشر للعامل ١ على ذلك البند.

بالطبع، وبعملية موازية، يمكن للعامل ٢ أيضًا التأثير على البند ليس فقط بشكل مباشر ولكن أيضًا بشكل غير مباشر من خلال علاقته بالعامل ١. نفس النوع من التأثير المباشر وغير المباشر للعوامل ينطبق على جميع البنود الأخرى. نتيجة لذلك، عند الحديث عن العلاقة بين بند وعامل فإنه عادةً ينبغي أن تكون هذه العلاقة مؤهلة بشكل صريح إما لإدراج أو استبعاد مثل هذه التأثيرات غير المباشرة. خلاف ذلك، يوجد غموض، وبالتالي ارتباك محتمل.

اختيار نوع التدوير:

كمسألة عملية، يجب أن يعتمد الاختيار بين الدوران المتعامد والدوران المائل على واحد أو أكثر من الاعتبارات. من بين هذه الاعتبارات كيف ينظر المرء إلى المفاهيم التي تمثلها العوامل. إذا كانت النظرية تشير بقوة إلى مفاهيم مترابطة، فربما يكون من المنطقي أن

يتبع النهج التحليــلي للعوامل (على وجه التحديد، الدوران) ذلك. وبالتالي، إذا كان تحليل العناصر المتعلقة بـ "الاجتهاد" و"الاعتمادية"، فإن الســماح للعوامل بالارتباط من شأنه أن يناســب إدراكنا بما تنطوي عليه هذه المفاهيم. عوضاً عن ذلك، قد تقترح النظرية عوامل متعامدة. على ســبيل المثال، قد تكون "الاعتمادية" و"المرح" أكثر اســتقلالية، وبالتالي قد تتطلــب حلاً متعامدًا. عندما لا تقدم النظرية دليــلاً قوياً، مثلما يمثل مقياس قيد التطوير مفاهيم لم تتم دراســتها من قبل، فإن حجم الارتباطات بين العوامل قد يكون بمثابة دليل. على وجه التحديد، يمكن تحديد الدوران المائل ودراسة قيم الارتباطات الناتجة بين العوامل. إذا كانــت قيم هذه الارتباطات صغيرة جدًا (على ســبيل المثال، أقل من ٠,١٥)، فقد يختار محلــل البيانــات الدوران المتعامد. هذا يقدم حلاً وســطاً لتقريب الهيكل البســيط ولكن ينتج عنه نموذج أبســط. على ســبيل المثال، قد تُظهر بعض البنود تشبعات ثانوية (بمعنى، تشــبعات على عامل آخر غير تلك التي تتشبع عليها بقوة أكبر) والتي تزداد بشكل طفيف بالنسبة للحل المائل ولكنها لا تزال صغيرة بشكل كافٍ لربط كل بند بشكل غامض مع عامل واحد فقط. وبالتالي، قد تكون تشبعات بند معين على ثلاثة عوامل تم تدويرها بشكل مائل هي ٠,٧٨ و٠,١٦ و٠,٠٥. عند اختيار الحل المتعامد، قد تكون التشبعات هي ٠,٧٧ و٠,١٩ و٠,١١. على الرغم من أن النموذج الثاني يبتعد أكثر بقليل من الأول عن الهيكل البســيط، إلا أنه لا يزال من الممكن ربط البند المعني بشكل لا لبس فيه بالعامل الأول. وبالتالي، تمت التضحية بالقليل في سبيل اختيار النموذج الأبسط (أي المتعامد) في هذه الحالة. إذا كانت العوامل مرتبطة بدرجة أكبر، فإن اختيار الحل المائل قد يؤدي إلى تحســن كبير في التقريب للهيكل البسيط. على سبيل المثال، قد يتضاءل تشبع ثانوي مقداره ٠,٤٠ تم الحصول عليه بتدوير متعامد إلى ٠,١٥ بحل مائل. لن يكون هذا هو الحال دائمًا، ومع ذلك، فإن دراســة الفرق بين طريقتي الدوران فقط يمكن أن تشير بشكل قاطع إلى أي مدى تختلف في تقريب الهيكل البسيط.

المســألة العمليــة الأخيرة تتعلق بحجم الارتباط بين عاملــين وكم يجب أن يكون حجم الارتباط قبل دمج العاملين في عامل واحد أكبر. لا توجد إجابة بسيطة على هذا السؤال لأن العلاقــات بين البنود والعوامل تحتاج أيضًا إلى النظر فيها. ومع ذلك، في بعض الحالات، قد يكشف الدوران المائل أنه حتى عندما يكون هناك عاملان يرتبطان ارتباطًا كبيرًا، فإن بعض

البنود لها تشـبعات كبـيرة على كليهما. في هذه الحالة، قد يكون من المنطقي اسـتخلاص عوامل أقل لمعرفة ما إذا كان العاملان المرتبطان إلى حد كبير يمكن دمجهما في عامل واحد. قد يكون، على سـبيل المثال، جيدًا أن تدعم بيانات الحياة الواقعية عاملاً واحداً يجمع بين البنود المتعلقة بـ "الاجتهاد" و"الاعتمادية" بدلاً من الفصل بينها.

النماذج العاملية الثنائية والهرمية:

تقريبًا، تفترض جميع أسـاليب تطوير المقاييس التي تمت مناقشـتها في هذا الكتاب أن البنـود المكونة للمقياس أحادية البعـد. التحليل العاملي هو الأداة المفضلة لتحديد البنود المرشـحة المحتملة التي ستناسـب هذا الافتراض. على الرغم من أن المجموعة الكاملة من البنـود التي يهتم بها الباحث قد تكون متعددة الأبعاد، إذا كان بإمكانه تحديد مجموعات فرعيـة أحادية البعد، فإنه عندئذٍ يمكن معاملة كل من هذه المجموعات الفرعية كمقياس منفصـل وأن العمل على تطوير أداة قياس (أو ربما بمزيـد من الدقة، مجموعة من أدوات القياس ذات الصلة) يمكن أن يستمر.

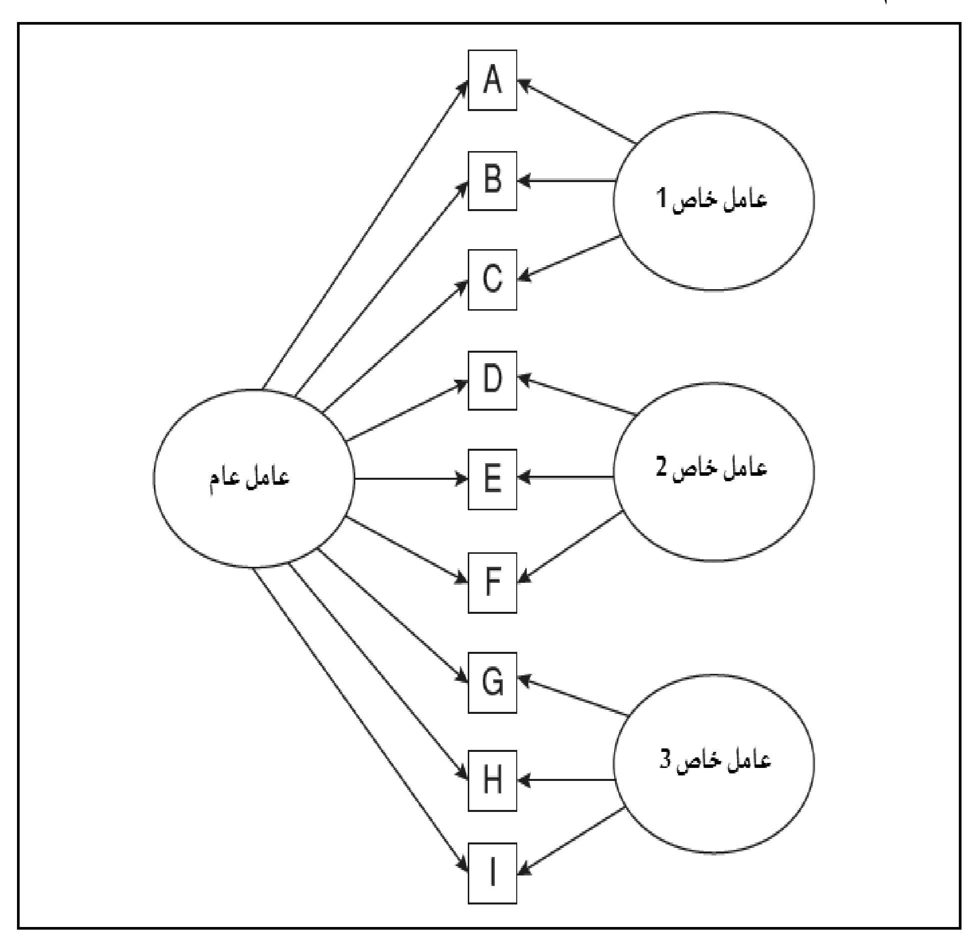
قبـل هذه النقطة، كانت مناقشـتنا للتحليل العاملي حول النمـوذج التحليلي للعوامل المتعـددة الأبعاد الأكثر شـيوعًا، أي نموذج قائم على هيكل بسـيط. تذكـر أنه تم وصف الهيكل البسيط بأنه موجود عندما يكون لكل بند تشبع قوي على عامل واحد فقط (والذي أشرنا إليه كعامل أسـاسي) وتشـبعات ضعيفة على جميع العوامل الأخـرى (أي الثانوية). وبإعادة الصياغة من منظور العوامل بدلاً من البنود، بالنسـبة لمفهوم الهيكل البسيط، يعد كل عامل محددًا قويًـا لبعض البنود (أي البنود التي نعتبرها مكونًا لذلك العامل) ومحددًا ضعيفًـا للبنـود المتبقية. نحن على وشـك النظر في بدائل لنماذج الهيكل البسـيط، ولكن نحتاج بعض التوضيح أولاً. النموذجان الموصوفان في الأقسام التالية هما، مرة أخرى، نماذج متعـددة الأبعاد (أي متعددة العوامل)، لكن العلاقات بـين البنود والعوامل لا تتوافق مع الهيكل البسـيط مثل النماذج السـابقة. مقارنةً بنهج الهيكل البسـيط، تمثل هذه النماذج طريقة مختلفة للتفكير في العوامل المتعلقة بالبنود. كما سنرى، بالنسبة لهذه النماذج، تتأثر البنود بما يمكن أن نسميه بالعامل الشامل global factor والعامل الموضعي local factor. وسـنقوم في النهاية بتحسـين هذه المصطلحات بحيث تتناسب مع كل نموذج من النماذج

التي نناقشها، ولكن هذه المصطلحات العامة ستخدم غرضنا في الوقت الحالي وستسمح لنا بالتمييز بين هذين النموذجين كثنائي عن تلك التي تطرقنا لها سابقًا.

لماذا نحتاج إلى نماذج تحليلية للعوامل غير تلك القائمة على الهيكل البسيط؟ بعض مجموعات البنود ببساطة لا تتوافق مع الهيكل البسيط. قد تُظهر عاملًا أوليًا (شاملاً) كبيرًا ولكنها تظهر أيضًا أدلة على مجموعات من البنود تشير إلى مكونات فرعية متماسكة ومتجانسة (عوامل موضعية). بمعنى أنه يبدو أن البنود لها عضوية مزدوجة، في العامل الأول الذي تشترك فيه جميعها وأيضًا في عوامل إضافية تشترك فيها مجموعات فرعية من البنود. السؤال الذي يطرح نفسه هو ما إذا كانت هذه البنود بأي حال من الأحوال أحادية البعد أم أنها متعددة الأبعاد لا محالة. بالمعنى الدقيق للكلمة، لا يمكن وصف البنود على أنها أحادية البعد استنادًا إلى الدليل بأن النموذج العاملي بسيط الهيكل سينتج. ومع ذلك، إذا تمكنًا بطريقة ما من وضع هذا العامل الأول الشامل جانباً، فإن علاقات البنود بالعوامل الموضعية قد تبدو إلى حدٍ كبير مثل المتغيرات الكامنة المعرّفة بمجموعات فرعية من البنود التي يبدو أن لها هيكلاً بسيطاً وبالتالي فهي أحادية البعد داخل المجموعات الفرعية. ولكن كيف يمكننا تجاوز مشكلة العضوية المزدوجة والتغلب على نواة الهيكل البسيط؟ النماذج التي تم وصفها هنا قد توفر الوسائل لذلك. وقد قدم (Reise, Moore, Haviland, 2010) ملخصًا ممتازًا لهذه النماذج، وتمت الاستفادة من رؤاهم بشكل كبير فيما سيأتي.

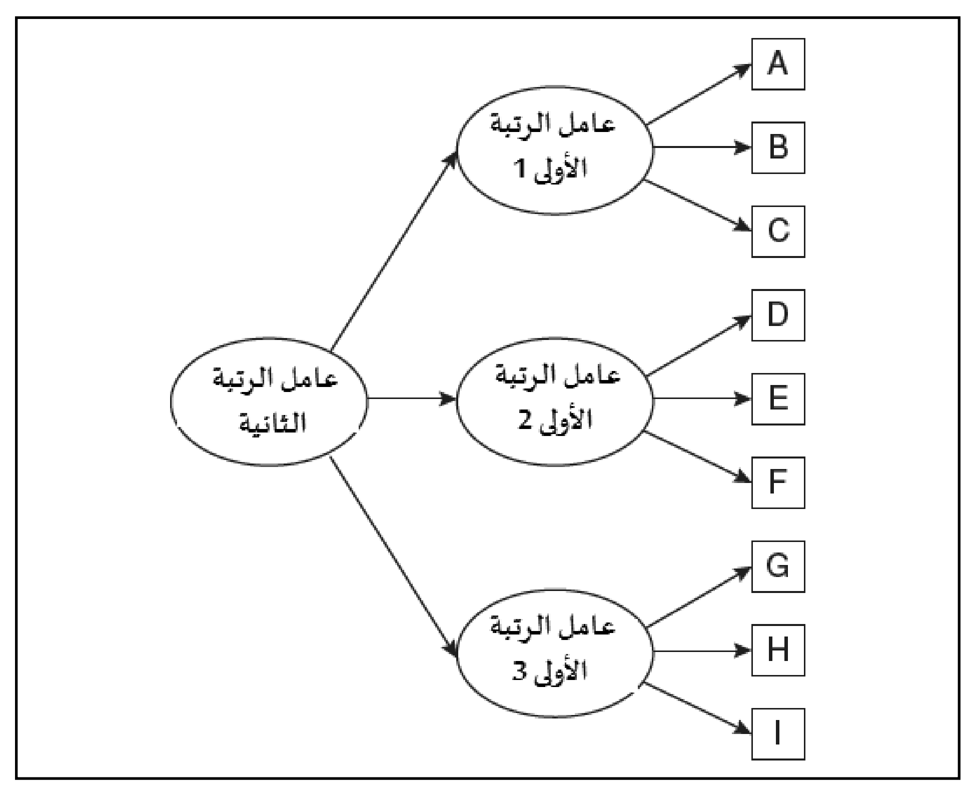
سننظر في نوعين من النماذج غير ذات الهيكل البسيط. نموذج ثنائي العوامل bifactor model ويحدد عاملاً عاماً (اسم للعامل الشامل في سياق النماذج ثنائية العوامل) يساهم في التباين في جميع البنود ويحدد أيضاً مجموعة من العوامل الخاصة specific factors غير المترابطة (مصطلح العوامل الموضعية داخل النموذج ثنائي العوامل) التي تساهم في التباين في مجموعات فرعية من البنود (انظر الشكل رقم ٦-١٤). وبالتالي، في المثال الموضح، يتلقى كل بند تباينًا من العامل العام بالإضافة إلى عامل واحد خاص فقط. باستخدام النموذج ثنائي العوامل، لدينا تشكيلة من المتغيرات التي تشبه أحادية البعد على مستوى السمات الخاصة على الرغم من وجود سمة عامة. إذا تم تطبيق النهج التحليلي العاملي التقليدي بسيط الهيكل على البيانات التي تتناسب مع نموذج ثاني العوامل، فإن البنود تميل إلى التشبع المتقاطع cross-load على العامل الأول (ملتقطة في المقام الأول

نفــس المتغير الكامن الذي يقــوم به العامل العام في نموذج ثنائي العوامل) وســمة أو أكثر من الســمات الخاصة. في هذا النهج، يؤدي حســاب العامل العام بشكل أساسي إلى تحديد التباين المشــترك بين جميع البنود. وما تبقى هي أكــثر مصادر التباين محدودية (العوامل الخاصة specific factors). قد تظهر البنود الموجودة داخل كل من هذه العوامل أحادية البعد الأساســية. بمعنى أن كل بند قد يتم تشــبعه بقوة على عامل واحد خاص فقط، وقد يحتــوي كل عامــل من العوامل الخاصة على بنود يمكن تفســيرها بســهولة على أنها تمثل مفاهيم بنائية constructs متمايزة ومتماسكة.



شكل (٦-١٤): نموذج ثنائي العوامل

من الواضح أن النموذج ثنائي العوامل مناســب لمجموعات البنود التي لها متغير كامن شــامل (أي، عام) وموضعي (أي، خاص) تمارس تأثــيرًا عليها. وينطبق هذا أيضًا على النهج التحليلي العاملي الآخر الذي قد يبدو مشــابهاً للنمــوذج ثنائي العوامل للوهلة الأولى، وهو النموذج الهرمــي hierarchical model. في النموذج الهرمي (انظر الشــكل رقم ٦-١٥)، تندرج مجموعة من العوامل الموضعية من الرتبة الأولى ضمن عاملٍ شــامل، أو أعلى رتبة، والتي تحدد المتغيرات الكامنة التي تمثلها عوامل الرتبة الأولى. على الرغم من أن المناقشــة التالية تفترض وجود نموذج ذي مستويين فقط (أي عوامل متعددة من الرتبة الأولى وعامل واحد من الرتبة الأعلى)، إلا أنه من الناحية النظرية يمكن أن يمتد التسلسل الهرمي ليشمل عوامل من الرتبة الثانية كأســباب لعوامل الرتبة الأولى، وعوامل من الرتبة الثالثة كأســباب لعوامل من الرتبة الثانية، وهلم جراً.



شكل (٦-١٥): النموذج الهرمي

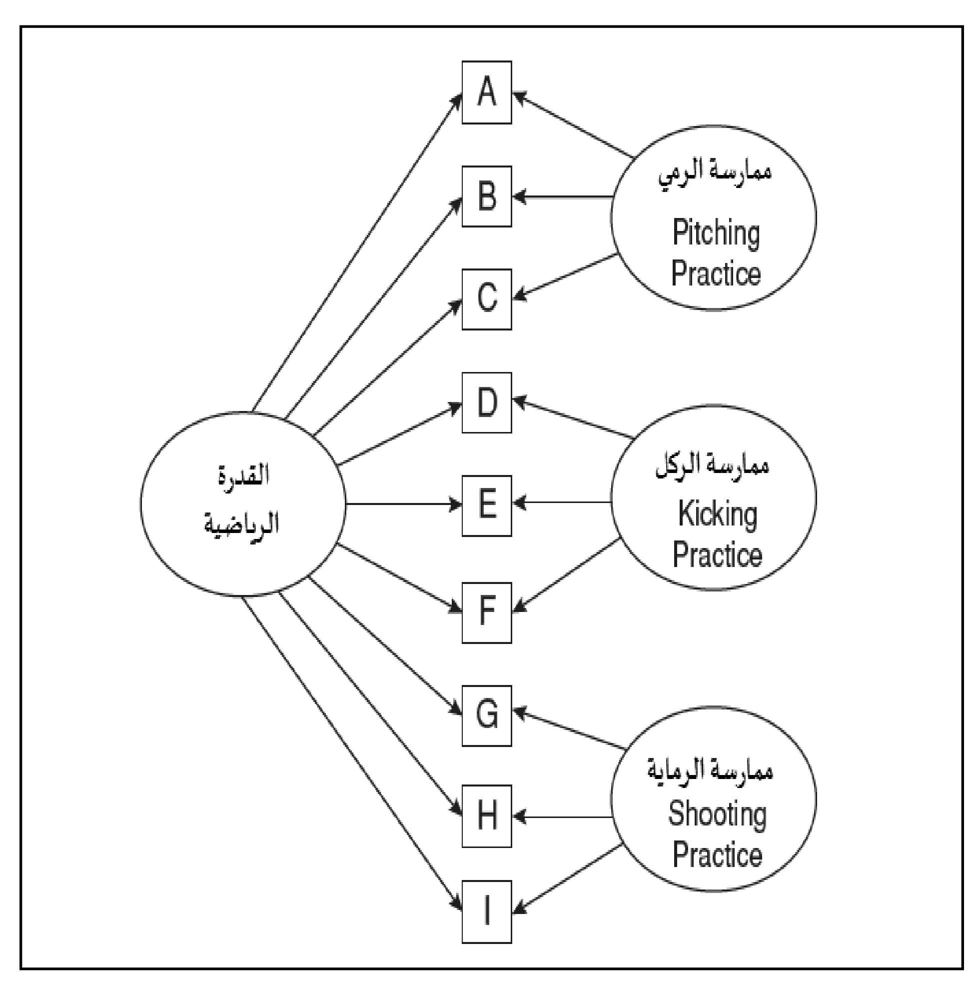
عــلى الرغم من أن النماذج الهرمية والنماذج ثنائية العوامل قد تبدو متشــابهة، حيث إن كليهــما ينطوي على عوامل شــاملة وموضعية، إلا أن هناك فرقًــا مفاهيميًّا مهمًا؛ إنهما يختلفان في مفهوم الســببية. في النموذج ثنائي العوامل، يتم تحديد قيمة كل بند بشــكل مبــاشر من خلال كل من العامــل الخاص والعامل العام، بينــما في النموذج الهرمي، يكون عامل الرتبة الأعلى ســببًا لعوامل الرتبة الأولى، والتــي هي بدورها تحدد قيم البنود. ليس لعامل الرتبة الأعلى أي تأثير سببي مباشر على قيم البنود، ولكن كل هذا التأثير يتم بوساطة عوامل الرتبة الأولى (Reise et al., 2010). وبالتالي، بينما تكون البنود هي مؤشرات للعامل العــام في النموذج ثنائي العوامل، فإنه في النموذج الهرمــي، تكون عوامل الرتبة الأولى هي المؤشرات (الكامنة) لعامل الرتبة الأعلى. بالنســبة للنمــوذج الهرمي من النوع الموضح في الشــكل رقم (٦-١٥)، فــإن التحليل الذي يؤدي إلى عامل الرتبة الأعلى ســيتضمن خطوتين مفاهيميتين: أولاً الحصول على الدرجات المناسبة من خلال التحليل العاملي لعوامل الرتبة الأولى (على ســبيل المثال، لكل عامل من الرتبة الأولى، متوســط قيمــة الدرجات التي يتم تشبعها عليه) ثم إجراء التحليل العاملي على درجات عوامل الرتبة الأولى تلك للحصول على تشبعاتها على عامل الرتبة الأعلى.

هنــاك مثالان مبســطان، واحد لــكل من هذه النــماذج التحليلية العامليــة الجديدة، ســيتم تناولهما مما قد يجعل هذه النماذج أكــثر وضوحًا. حيث يمكننا اعتبار ما قد يكون نموذجًا تحليليًا عاملياً مناســباً؛ أولاً لمجموعة من القدرات الرياضية ومن ثم بالنســبة للآراء الاجتماعية والسياسية.

قد يكون لدى الشخص مهارات رياضية عامة مثل القوة وخفة الحركة. بالإضافة إلى ذلك، قد يكون لدى هذا الشخص مهارات خاصة قائمة على الممارسة ذات صلة بالأنشطة الرياضية الفردية. قد تشمل هذه القدرة على رمي كرة سريعة عبر لوحة المنزل **home plate** في لعبة البيســبول، أو ركل كرة القدم من خلال الأعمدة القائمة **uprights** في كرة القدم الأمريكية، أو الإســقاط المباشر لكرة الســلة في الهدف من منتصف الملعب. هذه المهارات الأخيرة يتم تحديدهــا جزئيًــا من خلال القدرة الرياضية العامة ولكن يتــم تحديدها أيضًا إلى حدٍ كبير من خلال ممارســة المهارة الخاصة. وبالتالي، لا يمكن بالضرورة لأي شــخص لديه قوة جيدة

وخفــة في الحركة أن يرمي كرات سريعة بسرعة ٩٧ ميلاً في الســاعة التي كانت باســتمرار داخل الزاوية تماماً في منطقة ضرب كرة البيسبول. لذلك، إذا كان لدينا قدرٌ ما من مجموعة واسعة من المهارات الخاصة مثل قدرة الشخص على رمي ضربات الكرة السريعة، فقد يكون مــن الأفضل وضع نماذج للمؤشرات المقابلة لتلك المجموعة من المهارات باســتخدام نموذج ثنائي العوامل. هذا النموذج سيمثل كل مهارة كمهارة متأثرة بعامل القدرة الرياضية العام وعامل الممارســة الخــاص بالمهارات. وبالتالي، فإن قيم مــؤشر رمي ضربات الكرة السريعة ســتتأثر بشكل مشــترك بـ "القدرة الرياضية العامة" و"ممارســة الرمي". لاحظ أن سبب "ممارسة الرمي"، العامل الخاص في هذا المثال، ليس هو العامل العام، «القدرة الرياضية». بدلاً من ذلك، يســاهم هذان العاملان بشــكل مباشر في الاحترافية في رمي الكرة السريعة، ويوضح هذا الشــكل رقم (٦-١٦). البنود A وB وC ناتجة مباشرة عن كل من "ممارســة الرمــي" و"القدرة الرياضية". علاوة على ذلك، بعد الأخذ في الحســبان "القدرة الرياضية" العامة، فإن البنود التي تمثل "ممارســة الرمي" هي في الأساس أحادية البعد، وبالتالي يمكن أن تكون مناســبة للاندماج في مقياس. بشــكل أســاسي، فإن بنود "ممارســة الرمي" هي مجموعة أحادية البعد مندرجة ضمن النموذج الأوسع.

في حــالات أخرى، ســيكون النموذج الهرمي مناســبًا لبياناتنا أفضل مــن النموذج ثنائي العوامل. لنفترض الآن أن مشــكلة القياس التي نواجهها هي تحديد نموذج عاملي مناســب لمجموعة من مؤشرات وجهات النظر الاجتماعية والسياسية. قد تطلب البنود من المجيبين الإشــارة إلى شــعورهم تجاه سياســات الإنفاق الحكومي، مثل الدعم المالي لبرامج الفقر أو الــشركات الصغــيرة، وكيف ينظرون إلى القضايــا الاجتماعية للاختيار الشــخصي مثل زواج المثليين أو ملكية السلاح.

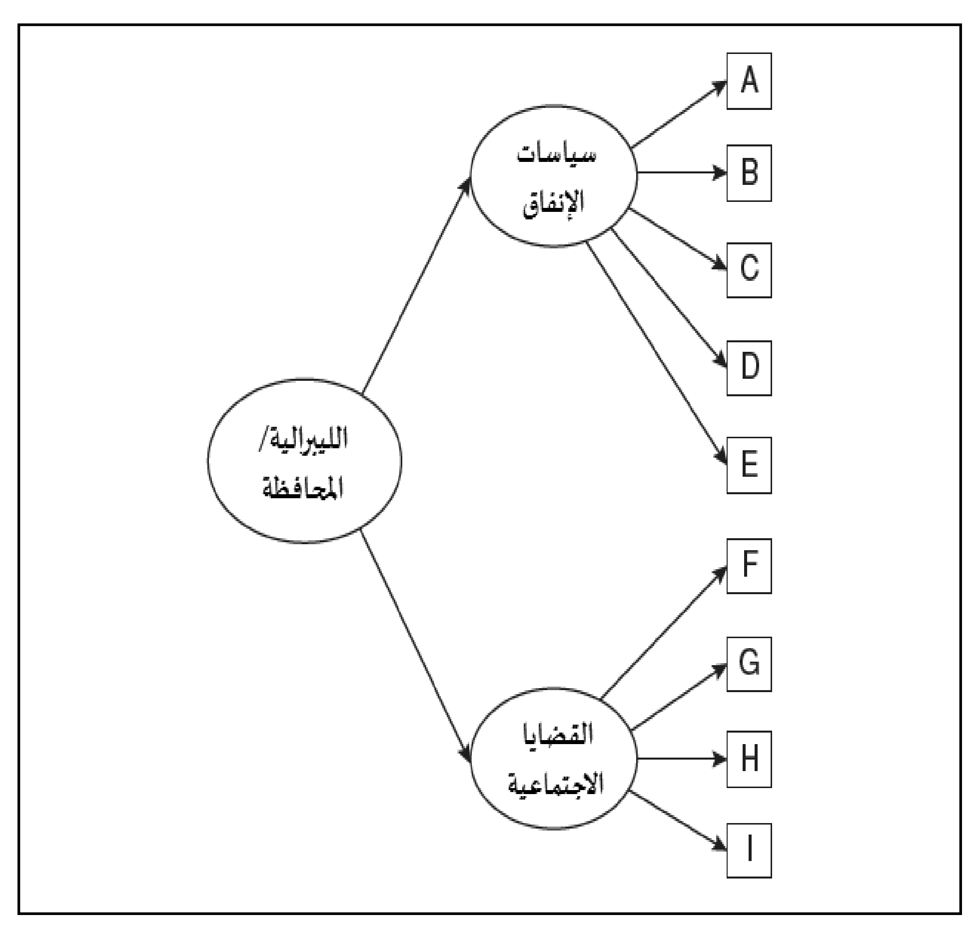


شكل (٦-١٦): نموذج ثنائي العوامل لـ "القدرة الرياضية"

كل فئة من فئات سياسة الإنفاق والقضايا الاجتماعية قد تحتوي على عدة بنود تمثل حالات خاصة لتلك المتغيرات. قد يكشف التحليل العاملي الأولي لهذه البنود المختلفة عن عاملين (من الرتبة الأولى) يشير محتواهما إلى أن أحدهما يستعرض وجهات النظر حول سياسات الإنفاق والآخر يستعرض المعتقدات حول الخيارات الشخصية. علاوة على ذلك، على الرغم من أن النتائج قد تشير إلى أن هذه العوامل متمايزة بوضوح، إلا أنها قد ترتبط ببعضها البعض. قد يستنتج الباحث أنه بالإضافة إلى العاملين الأساسيين الخاصين، هناك

عامل ســببي أعلى رتبة مثل "الليبرالية / المُحافِظة" التي تحدد آراء المجيبين بشــأن كل من سياســات الإنفــاق والقضايا الاجتماعيــة. وبالتالي، فإن "الليبراليــة / المحافظة" هي عامل محدد في وجهات النظر فيما يتعلق بكل من سياسات الإنفاق والقضايا الاجتماعية، ووجهتا النظر المحكمة هما المحددان الفوريان للاســتجابات للبنود الفردية. لاحظ أن "الليبرالية / المحافظة" ليست سببًا مباشرًا للاستجابات للبنود. على سبيل المثال، قد يكون لبعض الأفراد الذين يعتبرون محافظين أو ليبراليين بشكل عام آراء حول القضايا الخاصة قيد النظر والتي تتعــارض مع آراء معظم الليبراليين أو المحافظين. (مثال ملموس على ذلك هو أعضاء إدارة الرئيس الأمريكي ريجان المحافظة بشكل خاص، الذين لم يشاركوا آراء معظم المحافظين بعد محاولة اغتيال الرئيس فيما يتعلق بالســيطرة على الأســلحة). وعندما يكون الأمر كذلك، فإن العامل الثانوي البعد لن يكون لديه تأثير ســببي كبير على الاســتجابات لجميع البنود. أي إن التأثير الســببي للعامل الثانوي على اســتجابات البنود يتم بوساطة العامل الأساسي بدلاً من أن يكون مباشرًا. الترتيب المفاهيمي لهذه المتغيرات، مع تدفق السببية من عامل من الرتبة الثانية، إلى عامل من الرتبة الأولى، إلى البنود، يلائم نموذج تحليل العوامل الهرمي (انظر الشكل رقم ٦-١٧).

الغرض الأســاسي من اســتخدام نهج النموذج ثنائي العوامل أو النموذج الهرمي لتحليل العوامل هو الكشــف عن الأبعاد الكامنــة المتعددة التي تمتلكها بعــض البيانات وتحديد مجموعات البنود ذات البعد الأحادي بشــكل أساسي. النموذج ثنائي العوامل مفيد بشكل خاص في هذا الصدد للعديد من أنواع المتغيرات التي تمثل مؤشراتها مفهوماً بنائياً **construct** أكــثر عموميــة ومفاهيم بنائية متعددة أكثر خصوصية. يتيح نهــج النموذج ثنائي العوامل للفاحص بشــكل أساسي أن يضع العامل العام جانباً، ثم يدرس إلى أي مدى ينتج فعل ذلك مجموعات بنود أحادية البعد.



شكل (٦-١٧): النموذج الهرمي للآراء السياسية والاجتماعية

ومـع ذلـك، هذا الحـل لا يخلو مـن المضاعفات والقيود. على سـبيل المثال، سـتظل الاسـتجابات المشـاهدة للبنود تحتوي على تباين من كل من العوامل المشـتركة والعوامل الخاصة. وبالتالي، من المرجح أن تؤدي التقييمات البسـيطة للثبات التي لا تأخذ بالحسبان هيكليـة النمـوذج ثنائي العوامـل إلى المبالغة في تقدير الثبات. في الواقع، كما أشـار ريس وآخرون (Reise et al., 2010) أن كثيراً من الثبات الظاهر قد ينشأ عن العامل المشترك أكثر من العامل الخاص، حتى بالنسـبة للبنود التي تشـكل عاملًا خاصاً. وبالتالي، يجب اختيار تقديرات الثبات التي تعكس بدقة هيكل النموذج ثنائي العوامل للبيانات.

أيضاً، كما لاحظ ريس وآخرون (٢٠١٠)، أن الحزم الإحصائية التقليدية للتحليل العاملي المعــدة للنماذج بســيطة الهيكل تؤدي عملاً ضعيفاً في تحديــد وجود عامل عام مصحوب بعوامــل خاصة. وبالتالي، يجب على الباحثين اســتخدام برنامج مخصــص صراحة لتحليل النمــوذج ثنائي العوامل (مثــل الحزمة Psych ضمن بيئة تحليل البيانات R) لاستكشــاف البيانات التي تناســب هــذا النموذج (انظــر Reise et al., 2010، للحصول على مزيد من التفاصيل حول حزم البرامج المناســبة). يمكن أن تكــون هذه التحليلات معقدة إلى حدٍ ما وتســتفيد من خبرة الباحث وحكمه. العديد من الباحثين غير المختصين في القياس النفسي قد يختار بحكمة عدم محاولة إجراء تحليلات ثنائية العوامل دون توجيهات إرشــادية من الخــبراء. ومــع ذلك، من المفيد لجميع الباحثين المشــاركين في القياس أن يكونوا على دراية بهذه الامتدادات من التحليل العاملي - وخاصة نموذج ثنائي العوامل - لأنها قد توفر بدائل مفيدة لنماذج الهيكل البسيط لتحقيق أحادية البعد الأساسية.

تفسير العوامل:

في المثال الذي يتضمن بنود «الوعي» و«الاعتمادية»، افترضنا أننا نعرف سلفًا بالضبط ما هي المتغيرات الكامنة. في كثير من الأحيان، ليس هذا هو الحال، وحينئذ سوف نعتمد على التحليل العاملي لإعطائنا دلالات بشأن طبيعة تلك المتغيرات الكامنة. ويتم ذلك عن طريق فحــص البنــود التي تمثل بقوة كل عامل (على ســبيل المثال، البنود التــي تحتوي على أكبر تشــبعات على عامل معين). البنود ذات التشبعات الأكبر هي تلك الأكثر تشابهًا مع المتغير الكامن (وبالتالي، ترتبط بقوة أكبر). لذلك، يمكن أن توفر نافذة للتعرف على طبيعة العامل المعني. يتم القيام بذلك بســهولة عندما يكون هناك العديد من البنود التي تطرأ بوضوح على أحد المتغيرات العامة ذات التشــبعات الكبيرة (على ســبيل المثال، أكبر من ٠,٦٥) على نفــس العامل. بالعودة إلى مثال الخصائص التي تعتــبر مهمة لدى زميل العمل، إذا كانت الخصائص "ذكي" و"لديه ذاكرة حديدية" و"متعلم بشكل جيد"، وربما بند أو بندين آخرين ذات الصلة بالقدرة الفكرية تتشبع كلها بشكل كبير على نفس العامل، مع عدم وجود بنود لها تشبعات كبيرة على هذا العامل، فإنه سيكون من السهل إلى حد ما استنتاج أن "الأهمية المنسوبة إلى العقل" أو وصف ما مكافئ كانت تسمية ملائمة لهذا العامل.

على الرغم من أن اختيار تسمية لعامل ما قد يبدو واضحًا في بعض الحالات، فإن تعيين اســم لا يختلف عن إثبــات الصدق validity. إذا كانت مجموعة البنود تســتمر في الأداء كما يوحي الاســم المعين، فإنه سيتحدد الصدق في النهاية. عندما تفسر العوامل تباينًا قليلًا نســبيًا ولديها بنود متعددة ومتباينة ظاهرياً وتتشبع على نحو مماثل، فإنه يجب أن يكون محلل العوامل حذرًا بشكل خاص في التفسير. إذا أسفر التحليل عن عامل واحد ببنود تبدو مختلفة، فمن الأفضل عدم أخذ هذا العامل على محمل الجد كمؤشر على متغير كامن.

هناك نقطة أخرى تســتحق أن نتذكرها في مرحلة التفســير وهــي أن التحليل العاملي يمكن أن يعثر على البنية أو الهيكل التي تأخذ بالحســبان العلاقات فقط بين البنود التي تم تحليلها - لن تكشــف بالضرورة عن طبيعة الظواهر في حــد ذاتها. لا يمكن للباحث الذي يحاول تحديد الأبعاد الأساسية للشخصية، على سبيل المثال، الحصول على عامل "الانفتاح" إذا لم يتم تضمين أي بنود متعلقة بالانفتاح.

في بعض الأحيان، يمكن أن يؤدي تضمين عبارة أو جملة معينة إلى حدوث مظهر خاطئ لعامــل ذي مغزى مفاهيمي. عندما تتم صياغة بعض العبارات بصيغة المتكلم (أنا، نحن،) والبعــض الآخــر بدون، على ســبيل المثال، فإن ذلــك قد يفسر نمط الارتباطــات التي تمت مشاهدتها. كتوضيح، خذ بعين الاعتبار البنود الافتراضية التالية:

١- أنا أحب التفاح.

٢- للبرتقال طعم جيد.

٣- أنا أفضل التفاح على بعض الفواكه.

٤- هناك العديد من الأشخاص الذين يحبون البرتقال.

٥- أنا أستمتع بالتفاح بين الحين والآخر.

٦- البرتقال عموماً رائحته لطيفة.

٧- أنا أجد هشاشة التفاح جذابة.

٨- يمكن أن يكون البرتقال الطازج حلوى لذيذة.

إذا كانت البنود الفردية يتم تشــبعها على عامل واحد والبنود الزوجية على عامل ثانٍ، فلن نعرف ما إذا كانت الصياغة «أنا» للبنود الفردية هي ســبب العاملين أو إذا كان الناس

يعــبرون عن مواقف متباينــة تجاه النوعين من الفاكهة المذكورة. كل التفســيرات معقولة ولكن مربكة. هذه هي الحالة التي قد نقارن أو لا نقارن بها التفاح بالبرتقال.

طريقة المكونات الرئيسية مقابل العوامل المشتركة:

هناك فئتان عريضتان من أســاليب تحليل البيانات التي يعتبرها بعض المؤلفين أساســاً نفــس الشيء لكن آخرين يرون أنها مختلفة اختلافًا جذريًــا، هما التحليل العاملي وتحليل المكونات الرئيســية. أحيانًا يســتخدم مصطلح التحليل العاملي لاحتضان كلتا التقنيتين وفي أحيــان أخرى لوصف أحدهما على عكــس الآخر. غالبًا ما تســتخدم مصطلحات العوامل المشتركة common factors والمكونات components كوسيلة أقل غموضًا للإشارة بشكل خاص إلى المركبات الناشــئة عن التحليل العاملي وتحليل المكونات الرئيســية، على التوالي. هناك أساس لتأكيد كل من التشابه والاختلاف بين هذه الأساليب.

تحليل المكونات الرئيســية تنتج واحداً أو أكثر مــن المتغيرات المركبة التي تلتقط الكثير مــن المعلومات الموجودة أصلاً في مجموعة أكبر مــن البنود. علاوة على ذلك، يتم تعريف المكونات على أنها مجاميع مرجحة للبنود الأصلية. أي إن المكونات الرئيسية هي تحويلات خطية للمتغيرات الأصلية. إنها تتجذر في البيانات الفعلية وتستمد من البنود الفعلية. إنها مجرد إعادة تنظيم للمعلومات في البنود الفعلية.

تحليــل العوامل المشــتركة ينتج أيضًا واحــداً أو أكثر من المتغــيرات المركبة التي تلتقط الكثــير من المعلومات الموجــودة أصلاً في مجموعة أكبر من البنــود. ومع ذلك، تمثل هذه المركبــات المتغيرات الافتراضية. ولأنها افتراضية، كل ما يمكننــا الحصول عليه هو تقديرات لهذه المتغيرات. العامل المشــترك هو مفهوم بنائي مثالي تخيلي الذي افتراضياً يسبب البنود المراد الإجابة عليها كما هي؛ ويتم الاســتدلال على طبيعة المفهوم البنائي من خلال دراســة كيفية تأثيره على بنود معينة.

هل هما نفس الشيء أم مختلفان؟

يوضح الوصف أعلاه بعض الاختلافات بين المكونات والعوامل. أحد هذه الاختلافات هو أن العوامل تمثل متغيرات افتراضية مثالية نقدرها، بينما المكونات هي أشــكال بديلة للبنود

الأصلية من خلال ربط معلوماتها. الفكرة وراء اســتخراج العوامل المشــتركة هي أنه يمكننا إزالة التباين من كل بند لا يشترك مع أي من البنود الأخرى. من منظور التحليل العاملي، كما كان الحال مع الثبات، فإن التباين غير المشــترك هو خطأ في الأساس. وبالتالي، فإن التركيبات التي توصلنا إليها في استخراج العوامل المشتركة هي تقديرات للمتغيرات الافتراضية، الخالية مــن الأخطاء. وبهــذا المعنى، تكون العوامل المشــتركة مثالية - إنها تقديــرات لما قد يبدو عليــه متغير خالٍ من الأخطاء يحدد مجموعة من البنــود. علاوة على ذلك، تحدد العوامل كيــف تتم الإجابة عــلى البنود، في حين يتم تعريف المكونات مــن خلال كيفية الإجابة على البنود. وبالتالي، في تحليل المكونات الرئيسية، تعتبر المكونات منتجات نهائية للبنود، وتحدد الدرجــات الفعليــة التي تم الحصول عليها مــن البنود طبيعة المكونــات. بينما، في تحليل العوامل المشتركة فإننا نستحضر مفهوم المتغير الافتراضي المثالي الذي هو سبب درجات البند. العامل هو تقدير لهذا المتغير الافتراضي ويمثل سببًا، وليس أثراً، لدرجات البنود.

ماذا عن أوجه التشابه؟ هناك العديد:

أولاً، الفرق الحســابي بين الاثنين ضئيل. تذكر، في تحليل العوامل المشــتركة، أن الهدف هو تقدير متغير مثالي خالٍ من الأخطاء. ولكن يجب علينا إنشاء هذا التقدير من البيانات الفعليــة. كما أشرنا، تعتمد الطرق التحليلية للعوامل بشــكل عــام على مصفوفة الارتباط التــي تمثــل جميع الارتباطات بين البنود التي يجب أن تحلل إلى عوامل. في الفصل الثالث، أشرت إلى أن جميع القيم غير القطرية في مصفوفة التغاير أو الارتباط لا تمثل سوى التباين المشترك، الشيوعي communal. كما أشرت آنذاك، أن مصفوفة الارتباط هي مجرد نسخة معيارية standardized من مصفوفــة التباين - التغاير variance-covariance matrix. الارتباطات نفســها هي تغايرات معيارية، والوحدات unities هي تباينات معيارية للبنود. كل تبايــن معياري لبند يمثل كافة التباين، المشــترك والخــاص unique، الذي يظهره البند. لإنشاء متغير مثالي خالٍ من الأخطاء، يجب إزالة الجزء الخاص من التباين الموجود في تباينات البنود على طول القطر الرئيسي لمصفوفة الارتباط. وبشكل أكثر تحديدًا، يجب استبدال كل وحدة unity بتقدير الشيوع communality، وهي قيمة أقل من ١,٠، تقارب فقط التباين المشــترك لمتغير معين مع المتغيرات الأخرى المدرجة في التحليل العاملي. على ســبيل المثال،

إذا قدرنا أن هناك متغيرًا معينًا يشـارك ٤٥٪ من إجمالي تباينه مع بنود أخرى في مصفوفة الارتباط، فسنحدد له تقديرًا مشتركًا (الشيوع) بقيمة ٠,٤٥ ونضع تلك القيمة في المصفوفة، مع استبدال الرقم ١,٠ الذي مثَّل التباين الكلي للبند. سنفعل ذلك لكل متغير، مع استبدال كل وحدة بتقدير الشـيوع؛ (غالبًا، يتم الحصول على تقديرات الشـيوع عن طريق انحدار المتغـير المعني على جميـع المتغيرات المتبقية وذلك للحصول عـلى مربع الارتباط المتعدد، R^2، الـذي يقـوم بدور التقدير). تنتـج عملية التصفية هذه مصفوفـة ارتباط معدلة يتم اسـتخدامها لاسـتخراج العوامل المشـتركة بدلاً من المكونات، كما هو مبين في الجدول رقم (٦-١). التعويض بتقديرات الشيوع عن الوحدات unities هو الفرق الحسابي الوحيد الذي يميز استخراج العوامل المشتركة عن استخراج المكونات الرئيسية.

جدول (٦-١): مصفوفتا الارتباط لتحليل المكونات الرئيسية والتحليل العاملي المشترك

1.0	.70	.83	.48	.65
.70	1.0	.65	.33	.18
.83	.65	1.0	.26	.23
.48	.33	.26	1.0	.30
.65	.18	.23	.30	1.0

.45	.70	.83	.48	.65
.70	.52	.65	.33	.18
.83	.65	.62	.26	.23
.48	.33	.26	.48	.30
.65	.18	.23	.30	.58

ملاحظة: مصفوفة الارتباط في الأعلى، والتي تُستخدم لتحليل المكونات الرئيسية، تحتفظ بالوحدات في القطر الرئيسي. بينما تحتوي مصفوفة الارتباط في الأسـفل، والتي تُسـتخدم للتحليل العاملي المشترك، على تقديرات الشيوع، بدلاً من الوحدة، على طول القطر الرئيسي.

ماذا عن قضية السـبب والأثر؟ نحصل على كل من العوامل والمكونات من خلال تحليل الدرجات على البنود المشاهدة. كما يوضح شرح تقديرات الشيوع، فإن العلاقات التجريبية بـين البنود هي في النهاية أسـاس العوامل المشـتركة. وهذا، بالطبع، صحيح أيضًا بالنسـبة للمكونات. لذلك، حسـابيًا، كلاهـما يرتكز على البيانات التجريبيـة. علاوة على ذلك، فإن معظم محللي البيانات يصورون كلاً من المكونات والعوامل المشـتركة كطرق لفهم المتغيرات الكامنـة لمجموعة من البنود. وهذا يعني أن كلاً من المكونـات والعوامل يُعتقد عادةً أنها تكشف سبب الدرجات المرصودة لمجموعة من البنود. في الواقع، يستخدم تحليل المكونات وتحليـل العوامل، بالتبادل؛ حيث إنه في معظم الحـالات التي تكون فيها البنود لديها شيء مشـترك ذي معنى، فإن الطرق المختلفة تدعم نفس الاستنتاجات. في العديد من برمجيات تحليل العوامل (مثل PROCFACTOR في الحزمة الإحصائية SAS) يكون تحليل المكونات الرئيسـية هـو الافتراضي. بمعنى، يتـم الاحتفاظ بالوحدات في مصفوفـة الارتباط ما لم يتم تحديد الشيوع. لذا، على الرغم من أن هناك أوجه تشابه واختلاف بين الاثنين، إلا أن التمييز بينهما غالبًا ما يتم تجاهله مع وجود عدد قليل إن وجد من الآثار السلبية.

ومع ذلك، فإن أحد الاختلافات المهمة، هو طبيعة التباين التي تفسرها المكونات مقابل العوامل. تأخذ المكونات في الحسبان الجزء المحدد من التباين الكلي بين المتغيرات الأصلية، في حين أن العوامل تأخذ في الاعتبار التباين المشـترك أو الشـائع. إن تقليل القيم القطرية لمصفوفة الارتباط، كما يفعل الشـخص عند اسـتخراج العوامل المشـتركة، يقلل من البسط والمقام للنسـبة التي تعبر عن نسـبة التباين. لكنه يقلل من المقام إلى درجة أكبر بسـبب الحسابات المحددة التي ينطوي عليها حساب الفروق ذات الصلة. ونتيجة لذلك، فإن نسبة التباين التي تفسرها مجموعة من المكونات والعوامل القابلة للمقارنة لن تكون متسـاوية أو مكافئة من الناحية النظرية. سـوف تفسر العوامل نسـبة أكبر من التباين الأكثر تقييدًا (أي التباين المشـترك)، بينما تفسر المكونات نسـبة أصغر من التباين الكلي. عند مناقشـة النتائج التحليلية للعوامل وتقديم تقرير عن نسبة التباين المُفسَّر بالعوامل، فإنه من الأهمية بمكان أن نكون واضحين بشـأن نوع التحليل (المكونات أم العوامل المشتركة)، وبالتالي، نوع التباين (المشترك أو الإجمالي) الذي يتم تفسيره.

هنـاك اختلاف آخـر جدير بالملاحظة وهو أنـه في بعض الحـزم الإحصائية، فإن بعض المخرجـات التـي يتم الحصول عليها من اسـتخراج العوامل المشـتركة سـتبدو بلا معنى.

في كلا النوعين من التحليل، سوف يتزايد المقدار التراكمي للتباين المفسر كلما تم استخراج العامل/المكون التالي. في العوامل المشـتركة، غالبًا ما تتجاوز هذه النسبة ١,٠ في مرحلة ما، وتسـتمر في الارتفاع مع إضافة العوامل التالية، وبعد ذلك، كما لو كان الأمر سـحراً، تعود القيمة إلى ١,٠ تماماً عند استخلاص العامل K^{th}. على الرغم من أن هذا يبدو غريباً، إلا أنه مجـرد نتيجة ملفقة لطريقة حسـاب ويمكن تجاهلها. إذا اسـتخدم محلل البيانات معايير معقولة لتحديد عدد العوامل التي يجب اسـتخراجها، فإن العدد المختار سوف يسبق عادةً النقطة في تسلسـل الاستخراج حيث تنشأ هذه الحالة الشاذة. ومع ذلك، فمن الممكن لهذا العدد المحدد من العوامل أن يفسر تقريبًا كل (أي ١٠٠٪) التباين المشترك بين البنود الأصلية.

التحليل العاملي التوكيدي Confirmatory Factor Analysis:

التفرع الآخر لأسـاليب تحليل العوامل يميز بين الطرق الاستكشـافية والتوكيدية. تشير هـذه المصطلحات في الأصل إلى نية محلل البيانات بدلاً من الطريقة الحسـابية. وبالتالي، يمكن اسـتخدام نفس التحليل على نفس المجموعة من البنود إما لتحديد بنيتها الأساسـية (الاستكشـافية) أو لتأكيـد نمط معين من العلاقـات المتوقعة على أسـاس نظرية أو نتائج تحليلية سابقة (توكيدية). مع التكرار المتزايد، تُستخدم هذه المصطلحات الآن للتمييز بين أنواع مختلفة من الأدوات التحليلية بدلاً من الأهداف البحثية المختلفة. عندما يسـتخدم الأشـخاص مصطلح التحليل العاملي التوكيدي، فإنهم يتحدثـون غالبًا عن طرق قائمة على نمذجـة المعادلـة البنائيـة (structural equation modeling (SEM. وعلى الرغم من أنه يجب اسـتخدام هذه الأساليب بالطريقة التوكيدية بدلاً من الاستكشافية، إلا أنه يمكن اسـتخدام الأساليب القياسية التحليلية للعوامل لأيٍّ منهما. وبالتالي، فإن كلمة "التوكيدية" لا تعني بالضرورة نظام SEM.

ومع ذلك، فإن الأساليب المستندة إلى نمذجة المعادلات البنائية SEM تقدم بعض الفوائد الحقيقية التي تتميز بها على الأسـاليب التقليدية لتحليل العوامل في بعض الحالات. وتبرز هذه الفوائد لأن نماذج SEM مرنة للغاية. الشروط التي يتم افتراضها بواسـطة أسـاليب تحليـل العوامل التقليدية، مثل اسـتقلالية حدود الخطأ للبنود عـن بعضها البعض، يمكن تغييرها بشـكل انتقائي في الأساليب المستندة إلى SEM. كذلك، فإن الطرق التقليدية تقيد محلل البيانات إما بالسماح للعوامل بالترابط مع بعضها البعض أو تتطلب أن تكون جميعها

مســتقلة عــن بعضها البعض. الأســاليب القائمة على SEM يمكنهــا أن تجمع بين العوامل المترابطة وغير المترابطة إذا كانت النظرية تشير إلى أن مثل هذا النموذج ينطبق.

يمكن للطرق القائمة على نمذجة المعادلات البنائية SEM أيضًا توفير معيار إحصائي لتقييم مدى ملاءمة البيانات الحقيقية للنموذج المحدد. إن استخدامها بحكمة وتروي، يجعلها من الممكن أن تكون أصلاً. ولكن في بعض الأحيان، يمكن أن يؤدي ذلك إلى المبالغة في استخراج العوامل. حيث يؤدي اســتخراج المزيد من العوامل في كثير من الأحيان إلى تحسين ملاءمة النموذج. إن تطبيق معيار إحصائي صارم قد يحجب حقيقة أن بعض العوامل ذات الأهمية الإحصائية قد تمثل نســبًا صغيرة غير مثيرة للاهتــمام من التباين. خاصة في المراحل المبكرة من تطوير الأداة، قد يكون هذا مخالفاً لأهداف الباحث الذي يهتم بإيجاد أصغر عدد من العوامل المشبعة بالمعلومات بدلاً من حساب أكبر قدر ممكن من التباين.

وميــزة أخرى تتوفر في الأســاليب المعتمدة على نمذجة المعــادلات البنائية SEM وهي الممارســة الشــائعة لاختبار النماذج البديلة ومقارنة مدى ملاءمتهــا للبيانات. مرة أخرى، إن اســتخدامها بحكمة، يجعلها من الممكن أن تكــون أداة قيمة. وعلى العكس من ذلك، إن اســتخدامها بإهمال، قــد تؤدي إلى الوصول إلى نمــوذج ذي مواصفات أقل منطقية من الناحية النظرية ولكنها تؤدي إلى نموذج أفضل ملاءمة من الناحية الإحصائية. على ســبيل المثــال، قــد تؤدي إزالة القيد المتمثــل في عدم ارتباط أخطاء البنود مــع بعضها البعض إلى ظهور قيم صغيرة جدًا للارتباطات، ولكن قد لا يزال النموذج يتفوق من الناحية الإحصائية عــلى نموذج مقيد. قــد يقرر أحد الباحثين تجاهل الارتباطــات الصغيرة لصالح نموذج أكثر بســاطة، بينما يتم إقناع آخر بمعيار إحصائي لرفض البديل الأكثر بســاطة. كمثال آخر، قد يكــون النموذج الذي يفصل بين اثنين من العوامل المتمايزة ولكن مترابطة للغاية (ربما مثل "الوعي" و"الاعتمادية") أفضل من النموذج الذي يربط بين الاثنين. إذا كان الارتباط بينهما عالياً جداً، فقد يبدو قرار إبقائهم منفصلين تعسفيًا. على سبيل المثال، يمكن اعتبار الارتباط، لنقل ٠,٨٥، بين مؤشرين لنفس المفهوم البنائي construct دليلًا جيدًا على تكافؤهما. لكن النمــوذج الــذي يصف العوامل المنفصلــة التي ترتبط مع بعضها البعــض بمقدار ٠,٨٥ قد يناسب البيانات بشكل أفضل من النموذج الذي يجمع الاثنين في عامل واحد.

وليس المقصود من هذه التعليقات أن تشير إلى أن الطرق القائمة على نمذجة المعادلات البنائية SEM للتحليل العاملي التوكيدي هي ســيئة. قدم ظهور هذه الأساليب مساهمات

هائلــة لفهم مجموعة متنوعة من قضايــا القياس. ومع ذلك، أعتقد أن المرونة المتأصلة في هذه الأساليب تخلق فرصًا أكثر لاتخاذ قرارات رديئة، خاصةً عندما لا يكون محلل البيانات على دراية بهذه الأســاليب. مع اســتثناء محتمل لتحليل المكونات الرئيســية (حيث تكون العوامــل عبارة عــن مركبات خطية في البنود)، لا توجد طريقــة لتحليل العوامل تنتج حلاً صحيحًــا فريدًا. هذه الأســاليب تنتج فقط حلــولاً معقولة، والتي قد يكــون هناك الكثير منهــا. ليس هناك ما يضمــن أن النموذج الأكثر تعقيدًا الذي يتفــوق إحصائيًا على نموذج أبسط بديل هو انعكاس أكثر دقة للواقع؛ قد يكون أو لا يكون. مع كل الأساليب التحليلية للعوامــل، هناك حاجة إلى المنطق الســليم لاتخاذ أفضل القــرارات. التحليلات هي مجرد إرشادات لعملية صنع القرار وأدلة لدعم تلك القرارات. لا ينبغي لتلك التحليلات، في رأيي، أن تحل بالكامل محل صنع قرار الباحث. أيضاً، من المهم أن يتم وصف أســاس القرارات، إحصائية أو غير ذلك، بدقة في التقارير المنشورة عن التحليل العاملي التوكيدي.

ملاحظــة أخيرة حول هذا الموضوع: يعتبر الباحثون في بعض مجالات الاســتقصاء (مثل أبحاث الشــخصية) أن الحصول على نتائج متســقة من أســاليب تحليل العوامل التقليدية كدليــل تأكيــدي أقوى من إظهار نمــوذج جيد ملائم وفقًــا لمعايير إحصائية. على ســبيل المثال، ذكر Saucier and Goldberg (1996) أنه "نظرًا لأن التحليل العاملي الاستكشــافي يوفــر اختبارًا أكثر دقة للتكرار من التحليل العامــلي التوكيدي، فقد يكون التحليل العاملي الاستكشافي غالبًا مفضلاً على التحليل العاملي التوكيدي" (ص ٣٥). السبب هو أنه إذا كانت البيانات المأخوذة من عينات مختلفة من الأفراد في مناسبات مختلفة قد أسفرت عن نتائج تحليلية للعوامل متطابقة بشــكل أساسي باستخدام الأساليب الاستكشافية، فإن احتمال أن تكون هذه النتائج عبارة عن صدفة متكررة ضئيل للغاية. تذكر أنه في الأســاليب المستندة إلى SEM لهذا الموقف نفسه، يحدد محلل البيانات العلاقات المتوقعة بين المتغيرات ويحدد برنامج الكمبيوتر ما إذا كان يمكن مطابقة هذا النموذج مع البيانات التجريبية empirical. بمعنــى آخر، يُعطى الكمبيوتر تلميحًا كبيراً حول الطريقة التي يجب أن تصير عليها الأمور. في المقابل، فإن إعادة اكتشاف هيكلية سابقة للعوامل دون اللجوء إلى مثل هذه التلميحات، كما قد يحدث مع التحليلات الاستكشافية المتكررة، يمكن أن تكون مقنعة للغاية.

استخدام التحليل العاملي في تطوير المقاييس:

يجب أن يجعل المثال التالي بعض المفاهيم التي تمت مناقشتها في هذا الفصل أكثر واقعية. طور بعض الزملاء وأنا (DeVellis et al.,1993) مقياسًا لتقييم معتقدات الآباء بشأن من أو ماذا يؤثر على صحة أطفالهم. على الرغم من أن المقياس يحتوي على ٣٠ بنداً ويقيم العديد من جوانب هذه المعتقدات، إلا أنه بالنسبة لهذا العرض التقديمي، سأناقش فقط ١٢ بنداً من تلك البنود:

A- لدي القدرة على التأثير على رفاهية طفلي.

B- ما إذا كان طفلي يتجنب الإصابة هو مجرد حظ.

C- يلعب الحظ دورًا كبيرًا في تحديد مدى صحة طفلي.

D- يمكنني أن أفعل الكثير لحماية طفلي من الأذى.

E- يمكنني أن أفعل الكثير لحماية طفلي من المرض.

F- ما إذا كان طفلي يتجنب المرض هو مجرد مسألة حظ.

G- الأشياء التي أقوم بها في المنزل مع طفلي جزء مهم من رفاهيته.

H- سلامة طفلي تعتمد علي.

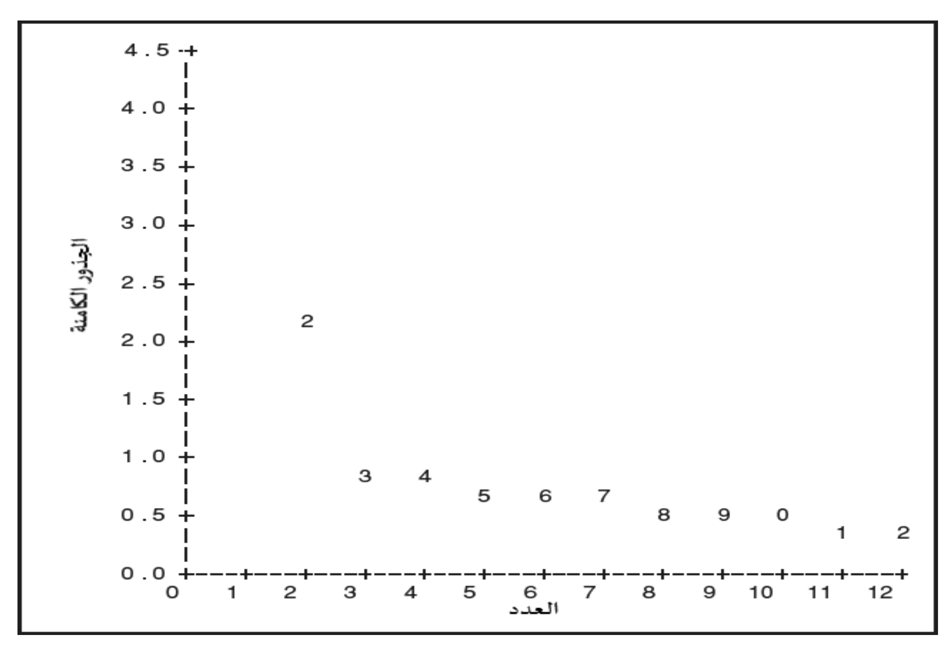
I- يمكنني أن أفعل الكثير لمساعدة طفلي على البقاء معافًى.

J- الصحة الجيدة لطفلي هي إلى حدٍ كبير مسألة حسن حظ.

K- يمكنني أن أفعل الكثير لمساعدة طفلي على أن يكون قويًا وصحيحاً.

L- سواء كان طفلي بصحة جيدة أو مريضاً هي مجرد مسألة قسمة ونصيب.

تم تطبيق هذه البنود على ٣٩٦ من الآباء والأمهات، وتم تحليل البيانات الناتجة باستخدام التحليل العاملي. كان الهدف الأول من تحليل العوامل هو تحديد عدد العوامل الكامنة وراء هذه البنود. وقد تم استخدام برنامج SAS لإجراء التحليل العاملي، وتنفيذ رسم الحصاة scree plot. يَظهر رسم الحصاة - مشابهاً للنوع المطبوع بواسطة SAS - أدناه (انظر الشكل رقم ٦-١٨). لاحظ أنه تم رسم ١٢ عاملًا (أي ما يصل إلى عدد البنود)؛ومع ذلك، يوجد ٢ من هذه العوامل على الجزء الأولي من الرسم والباقي يشكل الحصاة التي تمتد على طول قاع الرسم. هذا يشير بقوة إلى أن ٢ من العوامل مسؤولان عن الكثير من التباين بين البنود.



شكل (٦-١٨): رسم الحصاة الناتج عن التحليل العاملي للبنود المختارة

بعد تحديد عدد العوامل التي يجب الاحتفاظ بها، نقوم بإعادة تشغيل البرنامج الذي يحدد عاملين وتدويرهما بشكل متعامد (التباين الأعظم varimax). لو فشلنا في تقريب الهيكل البسيط، لقمنا بإجراء الدوران المائل oblique لتحسين الملاءمة بين البنود والعوامل. ومع ذلك، في هذه الحالة، أسفر الدوران المتعامد الأكثر بساطة عن مجموعات بنود ذات معنى لا لبس فيها وتشبعات قوية. وهذا يظهر جلياً في جدول تشبعات العوامل، كما هو موضح في الجدول رقم (٦-٢). يحتوي كل صف على تشبعات بند معين على العاملين. وقد قام أحد الخيارات المتاحة في برنامج SAS بإعادة ترتيب البنود الموجودة في الجدول بحيث يتم تجميع البنود ذات التشبعات العالية على كل عامل معًا.

في هذا الجدول، تم وضع خط تحت تشبعات البنود على العوامل التي تزيد عن ٠,٥٠. يتم تعريف كل عامل بالبنود التي يتم تشبعها عليه بشكل كبير (أي البنود التي تحتها خط). بالرجوع إلى محتوى تلك البنود، يمكن للمرء أن يميز طبيعة المتغير الكامن الذي يمثله كل عامل. في هذه الحالة، فإن جميع البنود التي يتم تشبعها بقوة على العامل ١ تهم الوالد باعتباره عاملاً مؤثراً على ما إذا كان الطفل يبقى آمنًا وصحيحاً.

جدول (٦-٢): تشبعات البنود على العاملين

	نمط العوامل المدارة *Rotated Factor Pattern*	
البنود	العامل1 *Factor 1*	العامل2 *Factor 2*
Item I	−.78612	−.22093
Item K	−.74807	−.18546
Item D	−.71880	−.02282
Item E	−.65897	−.15802
Item G	−.65814	−.01909
Item A	−.59749	−.14053
Item H	−.51857	−.07419
Item F	−.09218	−.82181
Item J	−.10873	−.78587
Item C	−.07773	−.75370
Item L	−.17298	−.73783
Item B	−.11609	−.63583

بينما تلك البنود التي تتشبع في المقام الأول على العامل ٢، تتعلق بتأثير الحظ أو القدر على صحة الطفل.

هاتان المجموعتان المتجانستان من البنود يمكن دراستهما بشكل أكثر. على سبيل المثال، يمكن حساب ألفا لكل مجموعة. حساب ألفا لمجموعات البنود هذه باستخدام برنامج SAS ينتج المعلومات الموضحة في الجدول رقم (٦-٣). كلا المقياسين يحتوي على معاملات ألفا للثبات بشكل مقبول. لاحظ أن الإجراء SAS CORR يحسب ألفا للبنود غير المعيارية والمعيارية. يكافئ حساب ألفا للبنود المعيارية استخدام صيغة ألفا المستندة إلى الارتباط. لكلا المقياسين، تسفر هاتان الطريقتان لحساب ألفا عن قيم متشابهة تمامًا. لاحظ أيضًا أنه لن تزيد قيمة ألفا بإسقاط أي من البنود لأي من المقياسين. قيم ألفا تقريبًا مرتفعة مثل تلك التي تم الحصول عليها للمقاييس الكاملة عند إسقاط بند واحد (على سبيل المثال، البند H من مقياس ١ والبند B من مقياس ٢). ومع ذلك، فإن الاحتفاظ بهذه البنود يوفر

بعــض الضمان الإضافي بأن الثبات لن ينخفض عن المســتويات المقبولة في عينة جديدة ولا تزيد أطوال المقاييس بشكل كبير.

جدول (٦-٣): معامل ألفا لجميع البنود ولجميع الـ 1 - k مجموعات مركبة من البنود لمجموعتين مختلفتين من البنود

ألفا	الارتباط بالمجموع	ألفا	الارتباط بالمجموع	البند المحذوف
معامل ألفا كرونباخ للبنود غير المعيارية: 0.796472؛ وللبنود المعيارية: 0.802006				
البنود المعيارية		**البنود غير المعيارية**		
.749666	.676138	.741489	.675583	ITEM I
.755695	.644648	.748916	.646645	ITEM K
.775939	.535924	.770329	.545751	ITEM E
.769222	.572530	.763252	.562833	ITEM D
.787007	.474390	.782509	.466433	ITEM G
.799245	.404512	.793925	.409650	ITEM H
.793003	.440404	.785718	.437088	ITEM A
معامل ألفا كرونباخ للبنود غير المعيارية:0.811162؛ وللبنود المعيارية: 0.811781				
البنود المعيارية		**البنود غير المعيارية**		
ألفا	**الارتباط بالمجموع**	**ألفا**	**الارتباط بالمجموع**	**البند المحذوف**
.749534	.682663	.748385	.684085	ITEM F
.776819	.594180	.775578	.596210	ITEM C
.763036	.639360	.762590	.636829	ITEM J
.777405	.592234	.776669	.593667	ITEM L
.806449	.493448	.806544	.491460	ITEM B

جميع المحاذير المطبقة على المقاييس بشــكلٍ عام في هــذه المرحلة من تطويرها قابلة للتطبيق على المقاييس المشتقة بأساليب التحليل العاملي. على سبيل المثال، من المهم جدًا تكرار ثبات المقاييس باستخدام عينة مستقلة. في الواقع، قد يكون من المفيد تكرار العملية التحليلية للعوامل بالكامل على عينة مســتقلة لإثبات أن النتائج التي تم الحصول عليها لم تكن حدثت مرة واحدة مصادفة.

حجم العينة:

إن احتمالية تكرار هيكلية العوامل هي جزئيًا على الأقل دالة في حجم العينة المستخدمة في التحليل الأصلي. بشكل عام، سيكون نمط العوامل الذي ينبثق من التحليل العاملي للعينة الكبيرة أكثر ثباتاً من ذلك الناتج من عينة أصغر. حتمًا، يطرح السؤال، "ما هو الحجم الكبير الكافي؟" هذا سؤال من الصعب الإجابة عليه (على سبيل المثال، MacCallum, Widaman, Zhang, & Hong, 1999). كــما هو الحــال مع العديد من الإجراءات الإحصائية الأخرى، ينبغي النظر إلى كل من نســبة عناصر subjects العينة لعــدد المتغيرات التي يتم تحليلها والعدد المطلق لعناصر العينة، لكن ثمة عوامل أخرى مثل تشبعات البنود تلعب أيضًا دوراً مهــمًا في ذلــك (MacCallum et al., 1999). كلما زاد عدد البنــود التي يجب أخذها في الحســبان وكلما زاد عــدد العوامل المتوقعة، ينبغي إدراج المزيد مــن العناصر في التحليل. مــن المغــري، بناءً على هــذه الحقيقة، البحث عــن معدل قياسي لعــدد العناصر إلى عدد البنــود. ومــع ذلك، كلما زاد حجــم العينة تدريجياً، يمكن أن يتناقــص معدل العناصر إلى البنــود. بالنســبة لتحليل عاملي لـ ٢٠ بنداً، من المحتمل أن تكــون ١٠٠ عنصر قليلة جدًا، ولكــن بالنســبة لتحليل عامــلي لـ ٩٠ بنداً، قد يكــون ٤٠٠ عنصر مناســبًا. يقترح كل من Tinsley and Tinsley (1987) معــدلاً مــن حوالي ٥ إلى ١٠ عناصر لكل بند وبما يصل إلى حــوالي ٣٠٠ عنصر. ويقترحان أنه عندما تكون العينــة بحجم ٣٠٠ عنصر، فإنه يمكن عدم التقيد بالنســبة بشــكل صارم. في نفس الورقة، استشــهدا بمجموعة أخرى من الإرشادات، منسوبة إلى Comrey (1973)، والتي تصنف عينة من ١٠٠ عنصر على أنها ضعيفة، و٢٠٠ عنصر على أنها متوســطة، و٣٠٠ عنصر على أنها جيدة، و٥٠٠ على أنها جيدة جدًا، و١٠٠٠ على أنها ممتازة. ذكر كومري (١٩٨٨) أن حجم العينة المكون من ٢٠٠ عنصر مناســب في معظــم حالات التحليل العاملي الاعتيادي التي لا تتضمن أكثر من ٤٠ بنداً. على الرغم من

أن علاقة حجم العينة بصلاحية الحلول التحليلية للعوامل أكثر تعقيدًا مما تشير إليه تلك القواعد القائمة على الخبرة، إلا أنه من الممكن أن تخدم الباحثين جيدًا في معظم الظروف.

بالتأكيد ليس من غير المألوف رؤية التحليلات العاملية المستخدمة في تطوير المقياس استنادًا إلى أحجام عينات أكثر تواضعًا (على سبيل المثال، ١٥٠ عنصراً). ومع ذلك، فإن النقطة الجيدة هي أن العينات الأكبر تزيد من تعميم الاستنتاجات التي تم التوصل إليها عن طريق التحليل العاملي. بالطبع، قد يكون تكرار حل التحليل العاملي على عينة منفصلة أفضل وسيلة لإثبات قابليتها للتعميم.

الخلاصة:

يعتبر التحليل العاملي أداة أساسية في تطوير المقاييس. حيث يسمح لمحلل البيانات بتحديد عدد العوامل التي تقوم عليها مجموعة من البنود مما يمكن تنفيذ الإجراءات مثل حساب ألفا كرونباخ بشكل صحيح. بالإضافة إلى ذلك، يمكن أن توفر لنا رؤية ثاقبة لطبيعة المتغيرات الكامنة التي تقوم عليها مجموعة البنود.

تمارين:

١- اشرح مفهوم الهيكل البسيط ومدى ارتباطه باختيار العناصر لإدراجها في المقياس.

٢- التحليل الموازي هو طريقة قائمة على الإحصاء لتحديد عدد العوامل أو المكونات الرئيسية التي يجب الاحتفاظ بها. صف بإيجاز كيف تحدد هذه الطريقة العوامل أو المكونات التي يجب الاحتفاظ بها.

٣- افترض أن مجموعة من المتغيرات يتم تحليلها مبدئيًا عن طريق استخراج العوامل المشتركة، وأن النتائج تكشف عن عامل أولي أساسي وعدة عوامل أصغر. علاوة على ذلك، يتم تشبع كل بند من البنود على كل من العامل الأول وواحد من العوامل الأصغر. يرغب الباحث في إنشاء مجموعة من المقاييس من هذه البنود، لكنه يشعر بالانزعاج من نمط التشبعات التي يكشف عنها هذا التحليل الأولي. ما الطريقة التحليلية للعوامل البديلة التي يمكن استخدامها لتحديد ما إذا كان يمكن تفسير البنود على أنها تنتمي إلى مجموعة من المجموعات أحادية البعد نسبياً؟

الفصل السابع
نظرة عامة على نظرية الاستجابة للبند (IRT)

نظرية الاســتجابة للبند (IRT) هي بديل لنظرية القياس الكلاســيكية، وتســمى أيضًا نظريــة الاختبار الكلاســيكية Classical Test Theory (CTT). وقد لاقت هذه النظرية اهتمامًــا متزايدًا في الســنوات الأخيرة، وغالبًا ما يتم تقديمهــا كبديل حديث ومتفوق على النظرية الكلاسيكية CTT (على سبيل المثال، De Boeck & Wilson, 2004; Embretson & Reise, 2010; Nering & Ostini, 2010; Reise & Revicki, 2015). كلتــا النظريتان تشــتركان في عدة خصائص أساســية. على ســبيل المثال، أحادية البعد للمقياس تعد شرطًا لــكل مــن نظريتَــي CTT وIRT. أي، يجب أن تشــترك البنود في متغير واحــد فقط إذا أريد دمجها في مقياس. بمعنى آخر، تحت هذين المنهجين، يجب أن تتقاســم البنود ســببًا مشتركًا واحدًا وبالتالي ترتبط ببعضها البعض. إذا كانت مجموعة من البنود متعددة الأبعاد (كما قد يكشــف التحليل العاملي)، فإنه يجب التعامل مع مجموعات البنود أحادية البعد المنفصلة بشــكل فردي. وهذا الأمر ينطبق في كلٍ من النهج الكلاسيكي CCT ونهج نظرية الاســتجابة للبند. لذلك، على ســبيل المثال، إذا شكلت مجموعة مكونة من ٥٠ بنداً خمسة عوامل منفصلة، فســيتم التعامل مع كل مجموعة من مجموعات البنود الخمســة بشكل منفصل، مما قد ينتج عنه خمسة مقاييس.

يُفرّق كلا الأســلوبين في القياس أيضًا بين تباين البند الذي ينشأ عن المستويات الفعلية للمتغــير الكامن والتباين الذي ينشــأ من الخطأ، لكن الطريقة التــي يتم بها التمييز بين الدرجــة الحقيقيــة true score والخطأ تختلف بين أســلوبي القياس. الفكرة الأساســية الكامنــة وراء نظرية القياس الكلاســيكية هي أن الدرجة المرصودة هي ببســاطة نتيجة للدرجة الحقيقية بالإضافة إلى الخطأ. مصدر هذا الخطأ لا تتم تجزئته إلى مصادر فرعية، مثــل الاختلاف في الوقت أو البيئة أو البنــود. بدلاً من ذلك، يتم جمع كل مصادر الخطأ في حد خطأ واحد.

بينــما تميز طرق نظرية الاســتجابة للبند الخطأ بشــكل أكثر دقــة، خاصة فيما يتعلق بخصائص البنود التي قد تؤثر على أدائها.

الهدف من نظرية الاستجابة للبند هو تمكين الباحث من تحديد خصائص معينة للبنود التي تكون مســتقلة عمن يجيب عليها. وهذا مماثل للقياس المادي الذي يمكن فيه تقييم سمة كائن ما (مثل الطول أو الوزن) دون النظر إلى الطبيعة المحددة للكائن. عشرون رطلاً، على ســبيل المثال، تعنــي نفس الشيء بغض النظر عما يتم وزنــه. وبالتالي، يعطي مقياس الوزن التقليدي معلومات حول خاصية محددة للأشياء (أي الوزن) بغض النظر عن طبيعة الكائن الذي يتم وزنه. تطمح طريقة نظرية الاستجابة للبند أن تعمل الشيء نفسه مع بنود الاســتبيان. الطرق الكلاسيكية تربط بعلاقة متأصلة بين أداة القياس والأشخاص الذين يتم قياســهم بينما لا تفعل طريقة نظرية الاستجابة للبند ذلك على الأقل من الناحية النظرية. على سبيل المثال، يتأثر ثبات المقياس وفقًا لنظرية القياس الكلاسيكية بالارتباطات بين البنود المكونة لهذا المقياس. فإذا كان الأفراد الذين تشكل بياناتهم أساسًا لتقييم الثبات متماثلين للغاية في مســتويات السمة الخاصة بهم، فإن مدى درجاتهم الحقيقية سيكون صغيراً. من نتائج قِصر المدى أن تكون الارتباطات بين البنود أصغر، وبالتالي ســيكون ثبات المقياس أقل بالنســبة للعينة التي يختلف فيها الأشــخاص على نطاق أوسع فيما يتعلق بالسمة المقاسة. وبالتــالي، فإن الثبات الذي يحصل عليــه الفرد وفقًا لنظرية القياس الكلاســيكية لا تتعلق فقط بأداء أداة القياس، ولكن في بعض الظروف، تتعلق أيضًا بخصائص العينة قيد الدراسة. أساليب IRT لا تبني تقييمات الثبات على أساس بيانات العينة بالطريقة نفسها التي تتبعها أســاليب القياس الكلاسيكية. بالطبع، في نهاية المطاف، لكلا الأسلوبين، يتم جمع معلومات البنود من الأفراد. وبالتالي، فإن تحقيق الميزة النظرية التي يمتلكها أسلوب نظرية الاستجابة للبند على أساليب القياس الكلاسيكية يعتمد على مدى تقييم البنود الموجودة تحت أسلوب نظرية الاستجابة للبند عبر عينات كبيرة غير متجانسة.

وصف جميع الاختلافات بين الأسلوب الكلاسيكي وأسلوب نظرية الاستجابة للبند خارج نطاق هذا الفصل. وللحصول على نظرة عامة عن أســلوب نظرية الاســتجابة للبند، يمكن الاطــلاع على Hambleton, Swaminathan, & Rogers (1991). بدلاً من محاولة إجراء مقارنة شــاملة، في الأقســام التالية، سأركز على ثلاثة أوجه تمييز رئيسية بين أساليب القياس

القائمة على نظرية القياس الكلاسيكية ونظرية الاستجابة للبند: (١) التركيز على البنود مقابل المقياس ككل، (٢) تحديد البنود ذات المستويات المحددة للسمة التي يتم قياسها، و(٣) التمثيل المرئي لخصائص البند والمقياس.

أول فرق مهم بين الأسلوبين هو أن نظرية الاستجابة للبند تولي اهتمامًا كبيرًا لخصائص البنود التي تشكل المقياس. في المقابل، تميل النظرية الكلاسيكية إلى التأكيد على خصائص المقياس ككل. كمثال على هذا الاختلاف، فكر في الثبات. عندما ناقشنا معامل ألفا كرونباخ في سياق القياس الكلاسيكي، لاحظنا أنه يمكن تعزيز ألفا إما عن طريق زيادة عدد البنود أو عن طريق تحسين متوسط الارتباط بين البنود. وبالتالي، فإن المزيد من البنود أو البنود الأفضل (بمعنى أن تكون مرتبطة بقوة أكبر مع المتغير الكامن) سوف يحسن الثبات. في نظرية القياس الكلاسيكية، غالبًا ما يزداد ثبات المقياس من خلال التكرار - إضافة المزيد من البنود. عادةً، يتم تعزيز الثبات في طرق IRT ليس بالتكرار ولكن (حيثما أمكن ذلك) من خلال تحديد بنود أفضل. وهذا يعني أن نهج IRT ينظر إلى الثبات بشكل أساسي من منظور البنود، في حين تنظر الأساليب الكلاسيكية إلى الثبات من منظور المقياس ككل. بينما يمكن للطرق الكلاسيكية، مثل فحص الارتباطات بين البنود والمجموع الكلي item-total correlation، تحديد كيفية مساهمة كل بند على حدة في الثبات الكلي للمقياس، إلا أن هذه العملية أقل ترسخًا في الأساليب الكلاسيكية منها في أساليب نظرية الاستجابة للبند لتطوير المقياس والتقييم.

يتمثل الاختلاف الثاني بين طريقتي القياس في أن نظرية الاستجابة للبند تفحص بشكل صريح مستوى السمة المقاسة الذي يؤثر بقوة على بند ما. قد يتم "ضبط" بنود مختلفة لمستويات مختلفة من السمة، وبالتالي، قد تكون حساسة للاختلافات على أجزاء مختلفة من السلسلة المتصلة الكاملة لتلك السمة. على سبيل المثال، ربما يقيس بندٌ مثل "أشعر أحيانًا بالحزن" مستوى أقل من الحزن أو الاكتئاب مقارنة ببند مثل "أشعر أن الحياة لا تستحق العيش". قد يميز البند الأول شخصًا نادرًا ما يشعر بمشاعر الحزن أو الاكتئاب عن الآخرين الذين يعانون من هذه المشاعر بشكل متكرر. وربما لا يكون مفيداً في التمييز بين الأشخاص الذين يشعرون بالحزن أحياناً عن أولئك الذين يشعرون بالحزن في الغالب. من ناحية أخرى، قد يميز البند الثاني فقط بين الأشخاص في الطرف العلوي من السلسلة

المتصلة للحزن أو الاكتئاب. وقد يفشـــل جميع الأشـــخاص على امتداد جزء كبير من هذه السلسلة في إقرار هذا البند، بينما الذين يعانون بشكل كبير من الحزن أو الاكتئاب يجيبون بشكل إيجابي. تحدد أساليب نظرية الاستجابة للبند هذين البندين على أنهما يمثلان نقاطاً مختلفة في سلســلة متصلة من الاكتئاب. من خلال تحديد البنود ذات الصلة بالنقاط على السلســلة المتصلة للمتغير الذي يتم قياسه، يمكن لطرائق نظرية الاستجابة للبند أن تساعد مطور الاختبار على تمثيل النطاق الكامل للسلسلة المتصلة للسمة التي يتم قياسها (الاكتئاب في هذا المثال) بالبنود المناســبة. بمعنى أنه يمكن تضمين البنود في مقياس يمثل مســتويات منخفضة ومعتدلة وعالية من الســمة بســهولة أكبر تحت نظرية الاستجابة للبند لأن هذه الطرق تربط البنود بشكل صريح بمستويات السمة التي تتعلق بها. في حين أن هناك طرق قياس كلاسيكية يمكن أن تحقق نتائج مماثلة (مثل فحص نسب المجيبين الذين يؤيدون كل بند وإجمالي الدرجات التي تصف المؤيدين وغير المؤيدين)، فإن هذه الطرق ليست خطوة عادية في طرق القياس الكلاســيكية. في المقابل، يعد تقييــم «صعوبة» البند جزءًا لا يتجزأ من معظم طرق نظرية الاســتجابة للبند. مدى قــوة ارتباط بند ما بالمتغير الكامن وموقع البند على طول السلسلة المتصلة للسمة، يكون لهما تأثير على الثبات. البنود المرتبطة بقوة بالمتغير الكامن سترتبط أيضاً ارتباطًا قويًا مع بعضها البعض، وبالتالي زيادة متوسط الارتباط الداخــلي بين البنود inter-item correlation وزيادة ثبات المقياس. علاوة على ذلك، من خلال تطوير بنود جيدة على طول النطاق الكامل للمتغير، يمكن للمرء أن يؤكد أن المقياس الناتج يمكن الاعتماد عليه عبر نطاق واسع من السمة التي يتم قياسها وليس فقط في أجزاء معينة منها. وبالتالي، يمكن أن يعزز كل من هذه الجوانب الشائعة والموجهة نحو البند في نظرية استجابة البند ثبات المقياس.

الاختلاف الثالث لنظرية الاســتجابة للبند بالنســبة لطرق القياس الكلاســيكية يرتبط ارتباطًا وثيقًا بالأولين. تســتخدم طرق نظرية الاســتجابة للبند على نطاق واســع الأشكال الرســومية كوســيلة لتمثيل خصائص البنود والمقاييس ككل. سننظر إلى بعض هذه الصور قريبًا، لكن أولاً، ســأناقش أصول بعض المصطلحات المســتخدمة عادةً لمناقشــة خصائص البنود تحت نظرية IRT وترتبط ارتباطًا وثيقًا بالأشــكال الرسومية التي تستخدمها نظرية الاستجابة للبند.

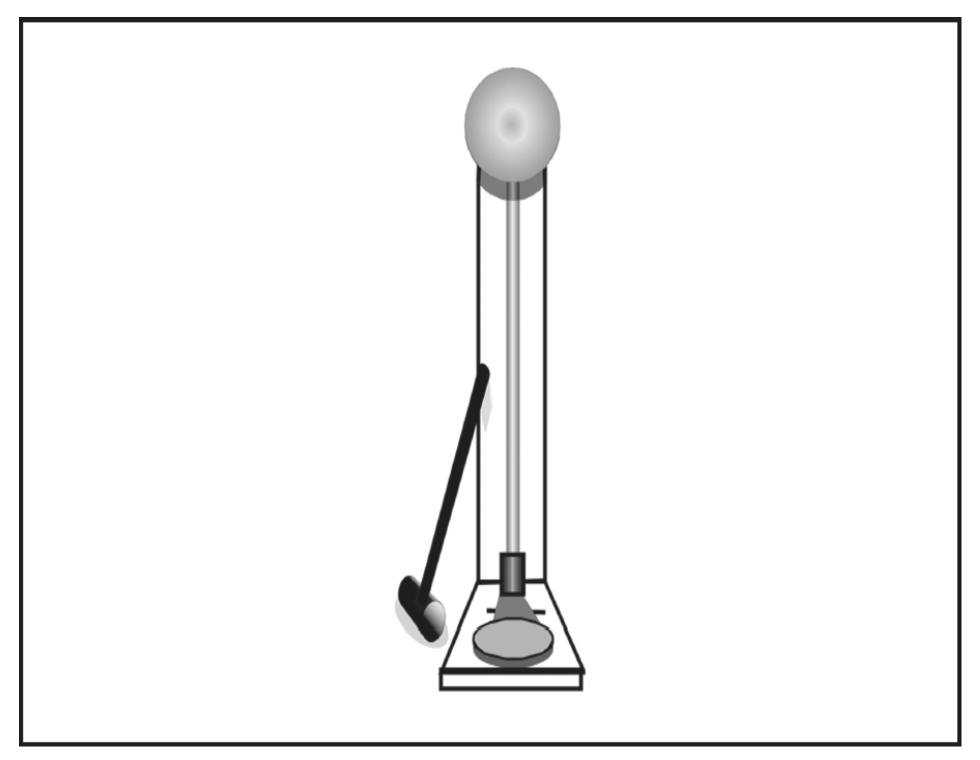
نظرًا لأن أسـلوب نظرية الاسـتجابة للبند نشأ في سـياق اختبار القدرات، فإن مفرداته vocabulary تحتـوي عـلى مصطلحات مرتبطة عـادةً بمنطقة المحتوى تلـك. أيضًا، نظرًا لتصنيـف البنود في اختبارات القدرة على أنها "صحيحة" أو "غير صحيحة" (على الرغم من أن تنسـيقها الأصلي قد يشتمل على أكثر من خيارين للاستجابة)، فإن التطبيقات التقليدية والأمثلة على نظرية الاسـتجابة للبنـد تتضمن بنوداً تأخذ واحدة من حالتين (على سـبيل المثال، «الاجتياز» مقابل «الفشـل»). من الأسـهل مناقشة نظرية الاسـتجابة للبند مبدئيًا للبنود من هذا النوع، على الرغم من أنه لا يوجد سـبب يمنع من توسـيع الطرق المنبثقة عـن النظرية (كما كانت بالفعل) لتشـمل بنوداً ذات صيغ اسـتجابة أخرى (مثل مقاييس Likert) تتطرق لمجالات المحتوى الأخرى. سـننظر في البنود من هذا النوع الأخير في وقت لاحق من الفصل.

نظرية الاسـتجابة للبنـد هي في الحقيقة مجموعة من النماذج بـدلاً من نظرية تحدد مجموعـة واحدة من الإجـراءات. إحدى الطرق المهمة التي تختلـف بها النماذج البديلة لنظريـة الاسـتجابة للبند هي عدد معلـمات البنود التي هي محل الاهتـمام. كان النهج المشـترك في السـنوات الأخيرة هو نموذج الثـلاث معلـمات three-parameter model، والذي، بشـكل غـير مفاجئ، يركز على ثلاثـة جوانب من أداء البند. وهـي صعوبة البند item difficulty وقدرتـه على التمييـز item discrimination والتخمين guessing - أو بشكل أعم، قابليته للإيجابيات الخاطئة susceptibility to false positives. أحد عناصر عائلـة IRT في وقت مبكر، وما زال يتمتع بشـعبية، هو نمذجـة راش Rasch modeling (على سبيل المثال، Rasch, 1960; Wright, 1999)، والتي تحدد فقط معامل الصعوبة.

صعوبة البند Item Difficulty:

عـلى الرغـم من أن مصطلـح "صعوبة البنـد" أتى - كما يتضح - من اختبـار القدرة، إلا أن المفاهيم التي يمثلها قابلة للتطبيق على نطاق أوسـع. تشـير صعوبة البند إلى مسـتوى السـمة المقاسـة المرتبطة بالانتقال من "الفشل" إلى "اجتياز" هذا البند. لقد شاهد معظمنا الأفلام القديمة التي تصور الكرنفالات أو المتنزهات التي تتميز ببعض العمل البطولي. "جهاز القياس" هو مسار عمودي ينتقل عبره الثُقل. في الجزء العلوي من المسار يكون هناك جرس.

في البداية، يقع الثقل في أسـفل المسـار عند نهاية لوح خشـبي يمثل قاعدة الجهاز. يضرب "المشـاركون" هذه القاعدة التي تقع مقابل الثقل باسـتخدام مطرقة كبيرة، وبالتالي يتحرك الثقل إلى أعلى على طول المسار. هدفهم هو دفع الثقل بقوة كافية لضرب الجرس. لأغراضنا، يمكننا أن نفكر في الجهاز بأكمله باعتباره "البند" (انظر الشكل رقم ٧-١).



شكل (٧-١): جهاز افتراضي لاختبار القوة التي يؤدي فيها ضرب القاعدة بالمطرقة بقوة كافية إلى رنين الجرس

تتمثل صعوبة البند في مقدار القوة التي يجب أن يمتلكها «المسـتجيب» (أو، على نحو أدق، القوة التي يجب أن يبذلها المستجيب) من أجل «اجتياز» البند (على سبيل المثال، دق الجرس). بوضوح، يمكن للمرء بناء بنود مختلفة بدرجات مختلفة من الصعوبة (على سبيل المثال، «أشـياء» أكثر صعوبة بمسارات أطول أو أثقال أكثر وزناً). ومع ذلك، يجب أن يكون من الممكن معايرة صعوبة جهاز معين يكون مسـتقلاً عن أي خاصية للشخص الذي يضرب بالمطرقة في تلك اللحظة.

نظرًا لأن هذا «الشيء» هو كائن مادي، فسـيكون من السـهل إلى حدٍ ما تحديد مقدار القوة اللازمة للتسـبب في رنين الجرس (مع التجاهل، في الوقت الحالي، تأثير ضرب القاعدة في مواقع مختلفة قليلاً عن نقطة الارتكاز). لذا، يُفترض أن يطلب مشـغل الكرنفال جهازًا بحجم ١٠ أرطال أو جهازًا يبلغ وزنه ١٠٠ رطل لتحقيق «معدل اجتياز» مرتفع أو منخفض حسـب الأشخاص الذين يلعبون اللعبة. كل من تلك الأوزان قد يكون مناسبًا بشكل خاص لمجموعـات مختلفة من العملاء (مثـل الأطفال الذين يحضرون معرضًا للمدرسـة مقابل البالغين الذين يحضرون معسكرًا رياضيًا).

يمكن للمرء أن يصف بنود اسـتبيان بطريقة مماثلة. بالنظر، على سبيل المثال، إلى بنود قياس الاكتئاب، كما فعلنا في وقت سابق. يمكن للمرء أن يصمم البند ليكون "سهلًا نسبيًا" (على سـبيل المثال، "أشـعر أحيانًا بالحزن") أو "صعبًا نسبيًا" (على سبيل المثال، "أشعر أن الحياة لا تستحق العيش"). في الحالة الأولى، ستكون هناك حاجة فقط إلى قدر متواضع من السـمة "الاكتئاب" "لاجتياز" (أي المصادقة) البند. لكن ألا يعتمد احتمال إصابة الشخص بهذا الشعور على من سئل؟ على سبيل المثال، إذا طرحنا هذا السؤال على الأشخاص المصابين سريرياً بالاكتئاب، فربما نجد أن نسـبة أكبر من تلك العينة تؤيد هذا البند أكثر مما لو قمنا بطرح هذا السؤال على الجمهور العام. الهدف من تحديد صعوبة البند هو تحديد مقدار السمة المطلوبة لاجتياز البند بالمعنى المطلق. إذا كان ذلك ممكنًا، فإن اجتياز الشخص لهذا البند له معنى ثابت فيما يتعلق بمسـتوى الاكتئاب، بصرف النظر عمن يكون هذا الشخص أو متوسط مستوى الاكتئاب في العينة قيد الدراسة. بمعنى آخر، يمكن وصف الشخص ليس فقط بالإشارة إلى عينة محددة ولكن من حيث المقياس المستقل عن أي عينة محددة.

تمييز البند Item Discrimination:

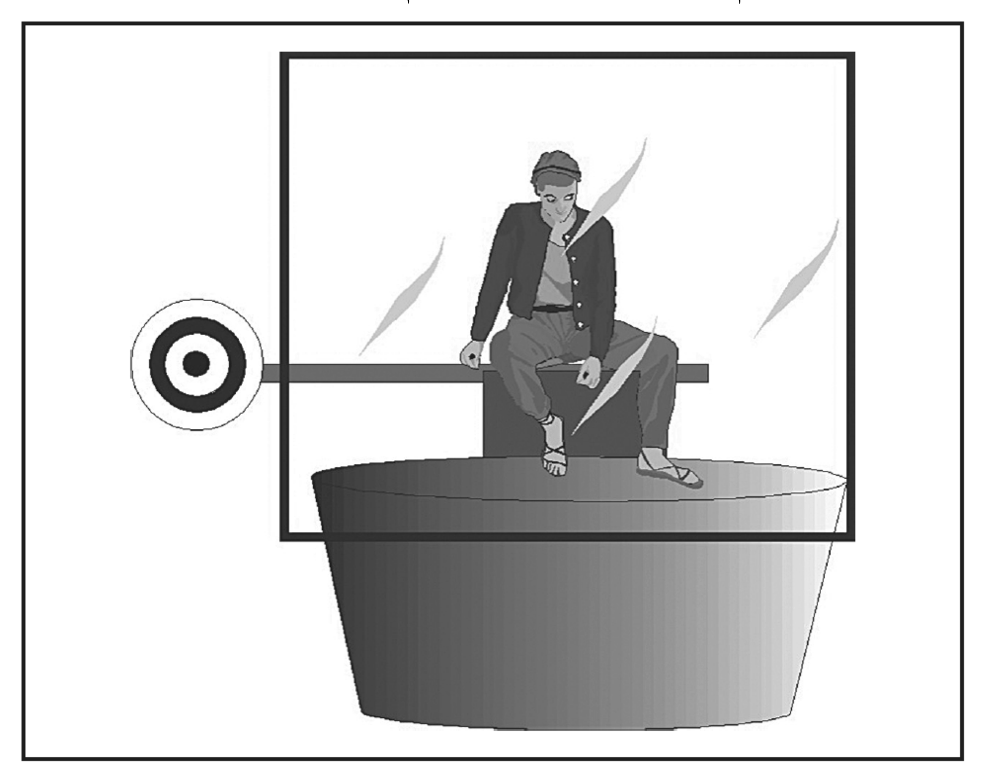
المعلمـة الثانية التي تتناولها نظرية الاسـتجابة للبند IRT هي الدرجة التي تصنف بها الاسـتجابة كـ "اجتياز" أو "فشـل" على أحد البنود بشكل لا لبس فيه. على نحو مختلف، كلما قل الغموض حول ما إذا كان الشـخص "اجتاز" حقًا أو "فشـل"، زاد التمييز في البند المعني. باسـتخدام تشـبيه الجرس الكرنفالي الذي تمت الإشارة إليه سابقاً، قد تكون هناك مناسـبات عندما يكون الثقل بالكاد يتصل بالجرس، مما يتسـبب في اختلاف المراقبين حول

مــا إذا كان الجــرس قــد رن بالفعل أم لا. قد يســمع البعض رنينًا خافتًا، بينما لا يســمع الآخرون شــيئًا على الإطلاق. ضمن نطــاق القوة التي تدفع الثقــل بحيث يلامس الجرس ولكنهــا لا تُنتــج ما يتفق عليــه الجميع على أنه صــوت رنين واضح، فــإن الجهاز في هذه الحالة يعطي معلومات غامضة. وبالنظر إلى هذا الغموض بطريقة أخرى، فإن نفس القوة المطبقة عدة مرات قد تؤدي إلى قيام المراقبين بتحديد أن الجرس قد رن في بعض المناسبات ولكنه لم يرن في مناسبات أخرى. وبالتالي فإن قوة أكبر إلى حدٍ ما سوف تنتج باستمرار رنيناً واضحاً لا لبس فيه، في حين أن قوة أضعف إلى حد ما سوف تنتج فشل (عدم رنين) لا لبس فيــه على الــدوام. ولكن هناك نطاق صغير من القوة التي يكــون فيها الجهاز غامضًا. قد يعمل جهاز بديل بشكل مختلف وتنتج عنه نتائج أقل غموضًا. على سبيل المثال، قد تؤدي ملامســة الثقل للجرس إلى فصل التيار عن الدائرة الكهربائية ومن ثم تحريك قواطع التيار الكهربائي مما يتسبب في إضاءة المصباح ويظل مُضاءً حتى تتم إعادة ضبطه. إذا تم تصميم هذا الجهاز بشكل جيد، فمن المحتمل أن يسفر عن نتائج متسقة على نطاق صغير إلى حدٍ ما من القوى، وبالتالي، ســوف يميز بشــكل أفضل من الجهاز القياسي. بدلاً من ذلك، فإن الجهاز الذي لم يكن له جرس على الإطلاق ولكنه بدلاً من ذلك طالب المراقبين برفع أيديهم إذا تجاوز الثقل خطًا محدد مسبقًا بجوار المسار الخاص به، من المحتمل أن ينتج عنه مزيد من الغموض، ومن ثم، فإن التمييز أقل. لذلك، فإن الجهاز أو البند الذي يميز جيدًا يحتوي عــلى جزء ضيق من نطاق ظاهرة الاهتمام التي تكون فيها النتائج غامضة. يحتوي الجهاز أو البند الأقل تمييزًا على مساحة أكبر من الغموض.

التخمين، أو الإيجابيات الزائفة Guessing, or False Positives:

المعلمــة الثالثة في نظرية الاســتجابة للبنــد هي التخمين، أو الإيجابيــات الزائفة. يعد مصطلح "التخمين"، مرة أخرى، إرثًا من أصول نظرية الاستجابة للبند في اختبار القدرة. في اختبار الاختيار من متعدد، يمكن للمســتجيب الذي لم يكن يعرف الإجابة الصحيحة فعليًا اجتيــاز البند عن طريق تخمين خيار الجواب الصحيح. هذا التخمين هو إيجابي زائف. أي أنه يعطي إشــارة إيجابية إلى أن المســتجيب يمتلك مســتوى معينًا من القدرة على الرغم مــن أنه لا يعرف بالفعل الإجابة الصحيحة عــلى البند المعني. وبالتالي، فإن اعتبار المعلمة

كإيجابيــة زائفة بدلاً من التخمين يســمح لنا بالتعميم بســهولة أكبر خارج ســياق اختبار القدرة. لعل تشبيه كرنفال آخر قد يوضح فكرة إيجابي زائف. ربما تكون قد رأيت أكشاك يجلس فيها شخص خلف حاجز واقي من نوع ما فوق خزان مياه على منصة متصلة بذراع تمتد إلى جانب مرسوم عليه الهدف (انظر الشكل رقم ٧-٢).



شكل (٧-٢): جهاز افتراضي لقياس دقة الرمي يؤدي فيه ضرب الهدف بالكرة إلى انهيار المنصة وسقوط الشخص الذي يجلس عليها في خزان الماء

يُطلب من المتســابقين إلقاء كرات البيســبول على الهدف، مما يــؤدي في حالة تعرضه لاصطدام الكرات إلى انهيار المنصة وسقوط الشخص في خزان المياه الذي يقع تحته. يمكننا التفكير في هذا الجهاز باعتباره «بنداً» يقيس دقة الرمي. يؤدي التسبب في سقوط الشخص مــن على المنصــة في خزان المياه إلى «اجتياز» هذا البنــد. (الآن، يجب أن تكون قادرًا على وصــف كيف يمكــن أن تؤدي الاختلافــات في الجهــاز إلى زيادة أو تقليــل صعوبة الجهاز

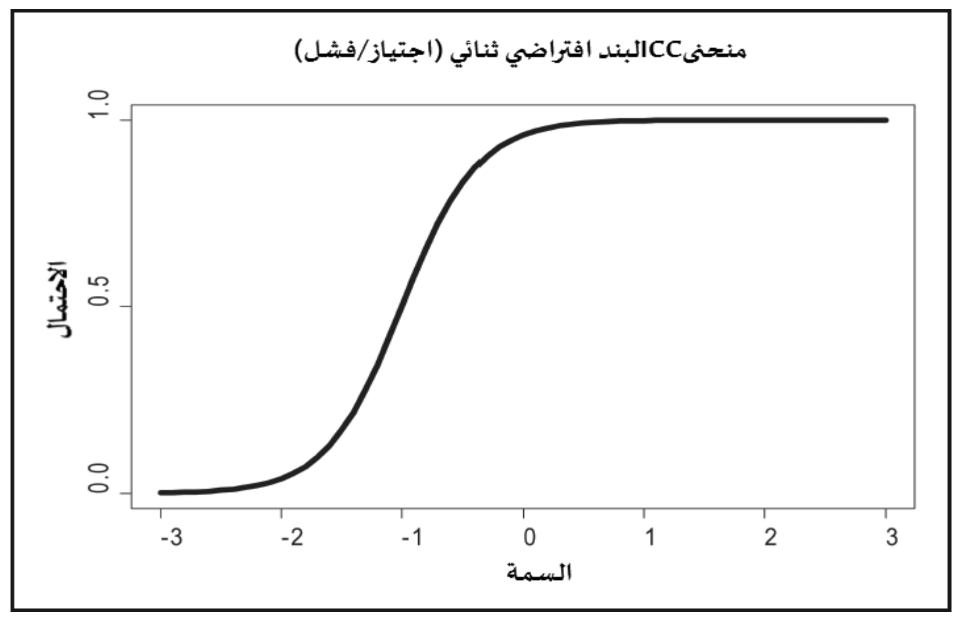
وتمييــزه). مع هذا الجهاز المحــدد، يمكن للمرء أن يتخيل كيف يمكن أن تحدث «إيجابيات زائفــة» - أي كيف يمكن للمشــارك بدون أي قدرة تقريبًا أن يســجل «اجتياز» عن طريق التســبب في ســقوط الشــخص الذي يجلس فوق الخزان. قد تكون إحدى الطرق هي أن «المشارك» يرمي بعنف ولكن الكرة تصيب الهدف فقط (يجب أن تذهب إلى مكانٍ ما، بعد كل شيء). أو، بدلاً من ذلك، قد يتعطل الجهاز وقد تنهار المنصة تلقائيًا. في هذه الحالات، يقوم اللاعب / المشارك «بالاجتياز» ليس بناءً على القدرة ولكن بناءً على بعض الظروف غير ذات الصلة. وبالتالي، من الممكن «اجتياز» اختبار الرمي هذا حتى لو كانت قدرة الشخص قليلة أو معدومة. كما ذُكر ســابقًا، في ســياق اختبار القدرات، تحــدث الإيجابيات الزائفة بشــكل شــائع نتيجة التخمين الناجح للإجابة الصحيحة على ســؤال ما على الرغم من عدم معرفة الإجابة حقًا. (في ســياقات القياس التي تكون فيها فرص التخمين أو الأنواع الأخرى من الإيجابيات الزائفة ضئيلة للغاية، مثل اســتخدام المقاييس لقياس الوزن، غالبًا ما يكون نموذج المعلمتين - أي الصعوبة والتمييز - كافياً).

كل مــن معلمات البند الثلاثة هذه - الصعوبة والتميز والإيجابيات الزائفة - لها علاقة واضحة إلى حدٍ ما بخطأ القياس. إذا كانت (أ) صعوبة بند ما غير مناســبة، أو (ب) تكون مســاحة الغموض بين الاجتياز والفشل كبيرة، أو (ج) يشير البند إلى وجود سمة ما حتى في حالــة غيابها، فإن البند يكون معرضاً للخطأ. تحدد نظرية الاســتجابة للبند هذه الجوانب الثلاثة لأداء بند ما، وبالتالي توفر وســيلة لاختيار البنود التي من المحتمل أن تعمل بشــكل جيد في سياق معين. الآن، لنرى كيف ترتبط هذه المعلمات الثلاث بالميزة الثالثة المهمة التي تميز أســاليب نظرية الاستجابة للبند عن الأساليب الكلاســيكية، وذلك من خلال استخدام الرسوم البيانية لتوضيح خصائص البند.

المنحنيات المميزة للبنود (ITEM-CHARACTERISTIC CURVES (ICC:

يمكــن تلخيص صعوبة البند وتمييزه وتوليد إيجابيات زائفة في شــكل المنحنى المميز للبند (ICC) الــذي يمثــل أداء البند بيانياً. عــادةً، يكون منحنى ICC على شــكل حرف S تقريبًا، وتكشف أجزاء مختلفة من المنحنى عن معلومات حول كل من المعلمات الثلاثة محل الاهتمام.

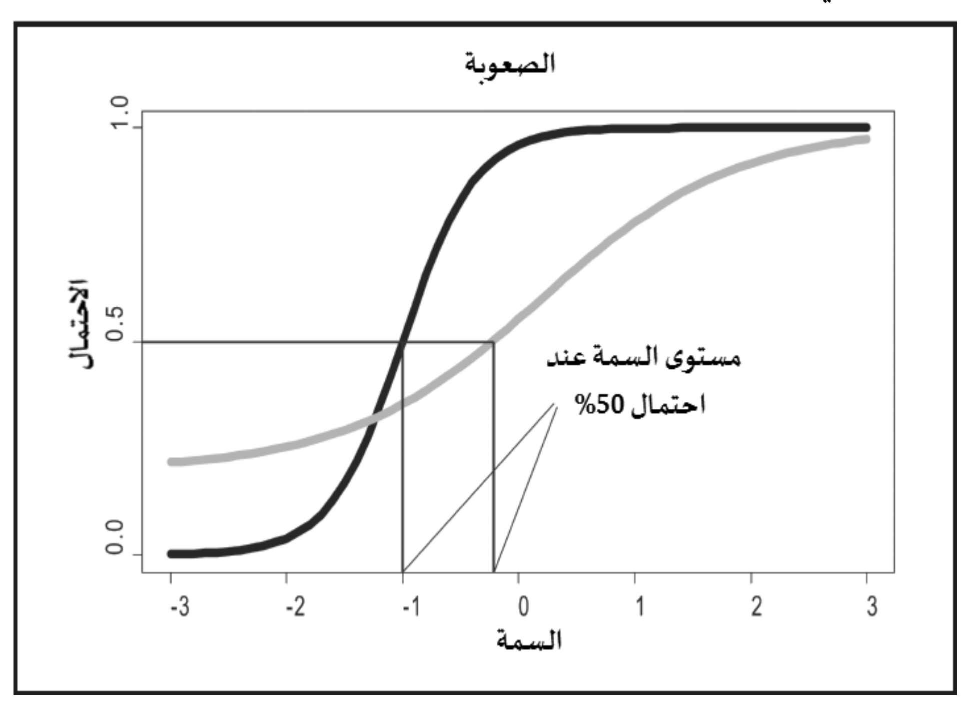
يوضح الشكل رقم (٧-٣) ما قد يبدو عليه شكل المنحنى المميز للبند ICC. حيث يمثل المحــور الأفقي X قوة الخاصية المميزة أو الســمة التي يتم قياســها (مثل المعرفة أو القوة أو الدقــة أو الاكتئاب أو الرغبــة الاجتماعية أو أي ظاهرة أخرى قابلة للقياس). بينما يمثل المحور الرأسي Y احتمال "اجتياز" البند المعني بناءً على نســب درجات الفشــل والاجتياز التــي يتم ملاحظتها. إن معرفة كيفية اســتخدام المنحنى المميــز للبند ICC لتقييم جودة البند يكون أسهل في الواقع إذا نظرنا إلى مخطط بياني يمثل بندين يمكننا مقارنتهما.



شكل (٧-٣): المنحنى الافتراضي المميز للبند (ICC) لبند ثنائي خيارات الإجابة (اجتياز/فشل)

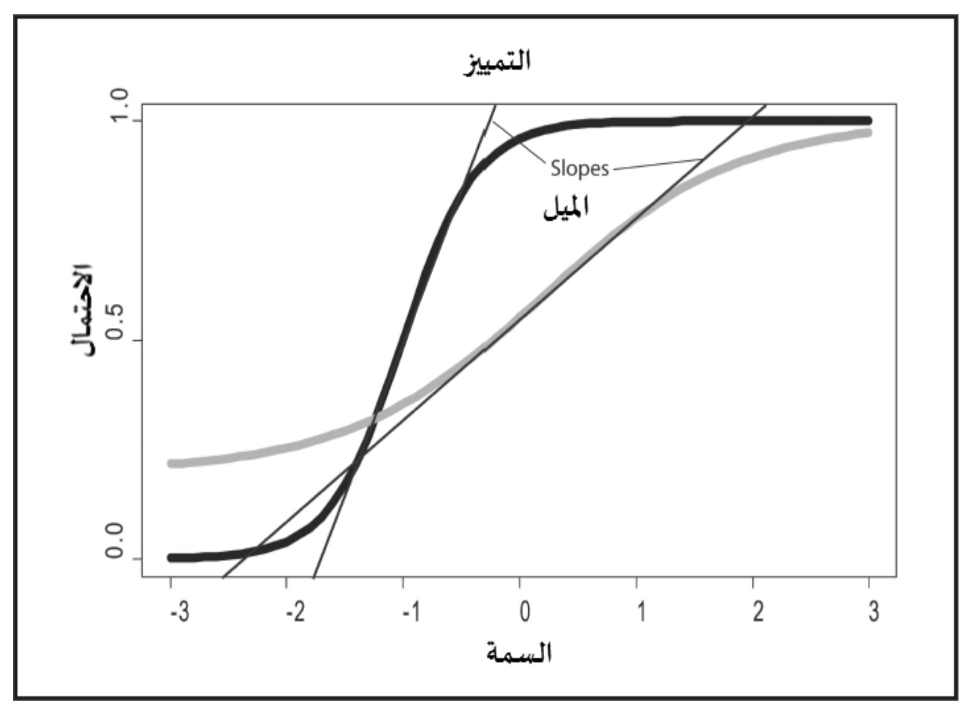
يوضح الشكل رقم (٧-٤) صعوبة البند من خلال عرض منحنيين. لاحظ أن النقاط التي يحقق فيها المنحنيان احتمالًا بنســبة ٥٠٪ في اجتياز بنودهما مختلفة. بالنســبة للمنحنى رمــادي اللون، تكون هــذه النقطة أبعد إلى اليمين. بمعنى، يجب أن يكون مقدار الســمة أعلى للفرد حتى يكون لديه فرصة بنســبة ٥٠٪ لاجتيــاز البند الذي يمثله المنحنى الرمادي مقارنةً بالبند الذي يمثله المنحنى أســود اللون. باســتخدام هــذا المعيار، يكون البند الذي يمثله المنحنى الرمادي أكثر صعوبة. الصعوبة في هذه الحالة ليست حكمًا ذاتيًا، وإنما وصفًا

واقعيًــا للنقطة عــلى المحور الأفقي X المقابلة لتقاطع المنحنــى لقيمة الاحتمال ٠,٥٠ على المحور الرأسي Y.



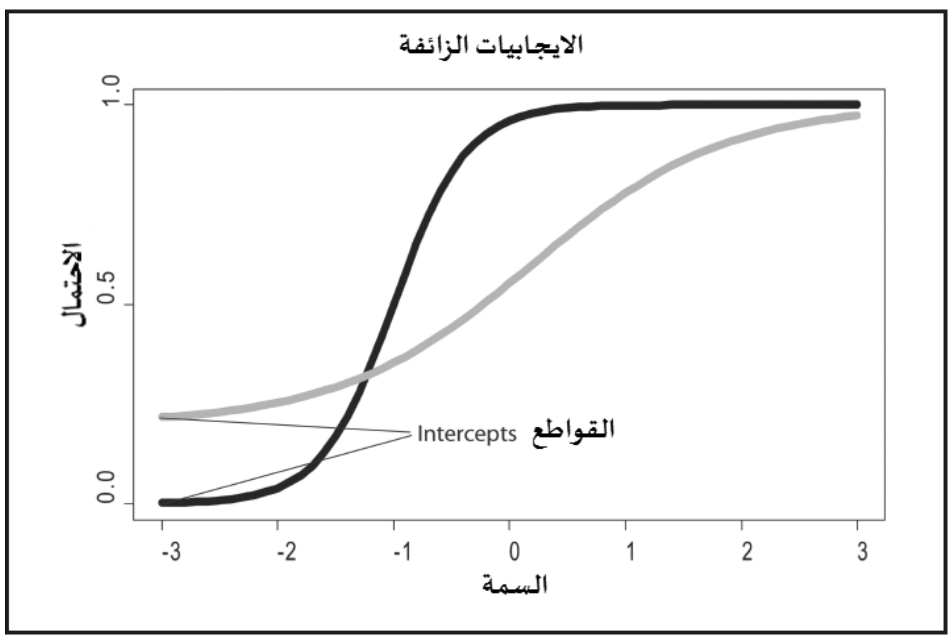
شكل (٧-٤): منحنيا ICC افتراضيان لبندين ثنائيين (اجتياز/فشل) مختلفين في درجة تحديد الصعوبة difficulty

يوضح الشــكل رقم (٧-٥) كيف نقوم بتقييم التمييز discrimination باستخدام نفس المنحنيان المميزان للبند ICCs. يحتوي البند المقابل للمنحنى أسود اللون على مَيل slope أكــثر انحدارًا عند نقطة الاجتياز ٥٠٪ من البند الذي يمثله المنحنى رمادي اللون. والنتيجة هي أن زيادة أقل في الســمة تعطي زيادة أكبر في الثقة بأن المســتجيب سيجتاز هذا البند مما هو عليه بالنسبة للبند الذي يمثله المنحنى الرمادي اللون. لذلك، يكشف المنحنى أسود اللون أن منطقة المحور الأفقي X التي تتوافق مع درجة غامضة أصغر من المنطقة المكافئة للبند الآخر. وبالتالي، فإن بند المنحنى أســود اللون يميز بين أولئك الذين يفشــلون والذين يجتازون بشكل أكثر فعالية من بند المنحنى رمادي اللون.



شكل (٧-٥): منحنيا ICC افتراضيان لبندين ثنائيين (اجتياز/فشل) مختلفين في تحديد درجة التمييز discrimination

أخيرًا، في الشـــكل رقم (٧-٦)، يمكننا أن ننظر إلى ميول البندين لإعطاء درجات الاجتياز حتى عندما تكون قدرة (أو أياً كانت السمة التي يتم قياسها) المستجيب في الأساس تساوي صفــراً. وكما كنـــت أنت قد خمنت، يتم تحديد ذلك من خـــلال النقطة التي يتقاطع فيها المنحنى المميز للبند ICC مع المحور الرأسي Y. بالنسبة للبند الممثل بالمنحنى أسود اللون، تكون قيمة القاطع intercept صفراً. وبالتالي، فإن احتمال قيام شخص ما باجتياز البند إذا كان يفتقر تمامًا إلى السمة المعنية سيكون صغيراً جدًا. بالنسبة للبند الممثل بالمنحنى رمادي اللون، فإن هناك احتمالاً كبيراً (حوالي ٢٠٪) بأن شـــخصًا لا يتمتع بالسمة أو القدرة المقاسة سيجتاز البند، وبالتالي، لا يمكن تمييزه، استنادًا إلى البند المعني، من شخص ذي قدرة عالية. (لاحـــظ أن هذا ســـيكون هو الحال مـــع بند متعدد الخيارات له خمـــس إجابات محتملة، إحداها كانت صحيحة). يوضح الرســـم التخطيطي في الشـــكل رقـــم (٧-٦) الاختلافات في قواطع intercepts المحور الرأسي Y للبندين - التي تعد الأساس لاستنتاج ذلك. مرة أخرى، البند الممثل بالمنحنى أسود اللون هو الأفضل أداءً.



شكل (٧-٦): منحنيا ICC افتراضيان لبندين ثنائيين (اجتياز/فشل) مختلفين في درجة تحديد الإيجابيات الزائفة false positives

من الناحية النظرية، يمكن اســتخدام نظرية الاســتجابة للبنــد IRT لتحديد المَعلَمات لكل بند من البنود. بعد ذلك، اعتمادًا على تفاصيل التطبيق المراد، يمكن اســتخدام البنود ذات الأداء الأمثــل لمعالجة مشــكلة القياس الموجودة. على ســبيل المثال، يمكن أخذ عينة من البنود "الســهلة" لتطبيقها على الأشــخاص الذين لديهم مستويات منخفضة نسبيًا من القدرة (السمة) المعنية وأخذ عينة من البنود الصعبة لأولئك الذين لديهم مستويات عالية من القدرة أو السمة المعنية. هذا يشبه بشكل مباشر استخدام جهاز جرس يزن ١٠ أرطال في معــرض أو كرنفال مخصص للأطفال وجهاز جرس يزن ١٠٠ رطل في معســكر للرياضيين البالغين. يمكن أن يؤدي اســتخدام البند الخطأ - مثل استخدام جهاز الجرس الخاطئ - إما إلى الإحباط (إذا كانت المهمة صعبة للغاية) أو نقص الدافع (إذا كانت المهمة سهلة للغاية). أيضًا، إذا كان المقياس الذي ســيتم توليفه ســيكون أساسًا لقرارات مهمة، فإن تقليل نطاق الغموض لكل بند واحتمال ظهور إيجابيات زائفة هي أيضًا احتمالات جذابة.

استـخدام الاختبار التكيفـي adaptive testing هو امتداد لهـذه الفكرة. في الاختبار التكيفي، البنود المطبقة على كل فرد يمكن تصميمها وفقًا لمسـتوى السمة عند ذلك الفرد. إن اجتياز بند سـابق أو الفشـل في ذلك، يمكن أن يرشـد إلى اختيار البند التالي، مع إعطاء الأفضلية لبند تم ضبطه على مسـتوى القدرة الذي تشير إليه الإجابات السابقة. تتم إدارة هذه العملية عادةً بواسـطة جهاز كمبيوتر وغالبًا ما يشـار إليها باسـم الاختبار التكيفي المحوسب computerized adaptive testing أو CAT. على الرغم من أن هذه المنهجية تتطلب مجموعة كبيرة من البنود التي تمثل مجموعة واسعة من مستويات القدرة، إلا أنها أداة قوية تتيحها طرق نظرية الاستجابة للبند IRT.

تتمتع طرق نظرية الاستجابة للبند IRT بميزة واضحة تتمثل في توجيه انتباهنا إلى ثلاثة جوانب (نسـخة المعلمات الثلاثة الشائعة حاليًا) لأداء بند ما. من خلال الأساليب المتجذرة في نظرية القياس الكلاسـيكية، قد نعرف (على سـبيل المثال، من أدائها في التحليل العاملي أو حسابات معامل ألفا) ما إذا كان بند ما يعمل بشكل جيد أو ضعيف ولكن قد لا يكون لدينا فهم واضح لطبيعة أي أوجه قصور لديه. بالمقابل، قد تساعدنا طرق نظرية الاستجابة للبند IRT على تقييم نقاط القوة والضعف في بند ما بشكل أكثر تحديدًا.

تطبيق نظرية الاستجابة للبند IRT على البنود متعددة الاستجابة:

تنطبق مقدمة نظرية الاستجابة للبند المقدمة أعلاه على نماذج ثلاثية المعلمات وتتعامل مع الاستجابات ثنائية التفرع، مثل صح أو خطأ. كما اقترح سابقًا، هناك العديد من نماذج نظرية الاسـتجابة للبند. وإنه من الشائع في العلوم الاجتماعية والسلوكية، أننا نتعامل مع أنواع من خيارات الاسـتجابات المرتبة ومتعددة المستويات والتي تمت مناقشتها بوجهٍ عام في الفصول السابقة.

بالنسبة للأدوات ذات خيارات الاستجابة متعددة المستويات، مثل مقاييس ليكرت Likert، يتم تطبيق نماذج نظرية الاسـتجابة للبند IRT المكيفة بشـكل خاص لتلك الأدوات. أحد أكثر هذه النماذج شـيوعًا هو نموذج الاسـتجابة المتدرج graded response model الذي طورته Fumiko Samejima (على سـبيل المثال، Samejima,1969) ويطبق بشكل شائع عن طريق برنامج David Thissen's Multilog software، (Thissen Chen & Bock, 2003). يوفر هذا النموذج معلومات حول كيفية ارتباط كل خيار من خيارات الاستجابة المتعددة بالقدرة.

فبالنســبة للبند الجيد، يجب أن يشــغل كل خيار اســتجابة جزءًا متميزًا أو أقل من القدرة المقاسة.

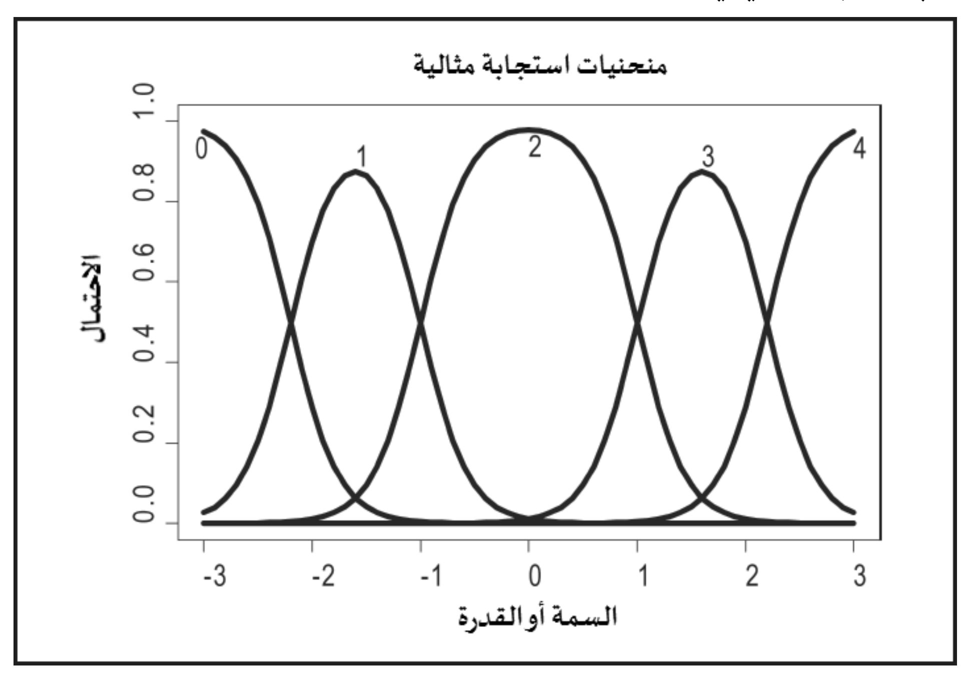
افترض، على سبيل المثال، بنداً مثل "أشعر بالدوار عندما أستيقظ للمرة الأولى في الصباح" وله خيارات الاســتجابة التاليــة: (٠) "أبدًا"، (١) "نــادرًا"، (٢) "بعض الوقت"، (٣) "معظم الوقــت" و(٤) "دائمًــا تقريبًا". يتوقع المــرء أن احتمال اختيار خيارات أعــلى تدريجيًا على المقياس يرتبط بمســتويات أعلى تدريجيًا من الدوار (الدوخة) الصباحي (أي مستويات أعلى تدريجيًا من الســمة محل التقييم). بطريقة مــا، يحل بند مثل هذا محل العديد من البنود ثنائية الاستجابة. لذا، فبدلاً من البنود الثنائية من النوع "نعم - لا" المنفصلة، مثل "لا أشعر بالدوار أبدًا عندما أستيقظ للمرة الأولى في الصباح"، "نادرًا ما أشعر بالدوار عندما أستيقظ في الصباح"، "في بعض الأحيان أعاني من الدوار عندما أســتيقظ في الصباح". وهكذا، يتم تقديم بند واحد له العديد من خيارات الاستجابة التي يمكن للشخص المستجيب الاختيار من بينها.

إذا كان البند يعمل بشكل جيد على مجتمع معين، فإن احتمالية اختيار خيار الاستجابة الأول (أي "أبدًا") يجب أن تكون مرتفعة للأشــخاص ذوي المســتويات المنخفضة للســمة (الدوخة) ويجب أن تتضاءل فعلياً إلى الصفر مع الزيادة في مســتوى السمة. لذلك، ينبغي أن يكــون المنحنى الذي يمثل احتمال اختيار خيار الاســتجابة الأول عالياً في أقصى اليســار (المنطقة المقابلة لمســتويات منخفضة من الدوخة في الصباح) ومنخفض عند أقصى اليمين. في الطــرف الآخر من المقياس، يجــب أن يكون احتمال اختيار خيار الاســتجابة الأخير (أي "دائمًا تقريبًا") صفريًا تقريبًا للأشخاص في المستويات المنخفضة إلى المتوسطة من السمة، مع التقدم والاقتراب من نسبة ١٠٠٪ من احتمالية الدوخة لــلأشخاص الذين يعانون من أعلى مســتويات الدوخة في الصباح. يجب أن يكون منحنى هذه الاستجابة منخفضًا للغاية عند النهايات المنخفضة من مســتويات الدوخة الصباحية، بحيث يصل إلى الحد الأقصى للقيمة عند أقصى اليمين من مســتويات الدوخة. وبالتالي، فــإن منحنيَي كلٍ من هذين الخيارين للاستجابة القصوى غير متناظرة nonsymmetrical، حيث إن أحد الطرفين منخفض للغاية والآخــر مرتفع جداً. وأيضًا، يجب وضع كلٍ من هذين المنحنيين على أقصى حد من النطاق المطابق لمســتوى السمة التي يمثلها خيار الاستجابة. على النقيض من ذلك، يجب أن تكون المنحنيات التي تقع بين خيارات الاستجابة متماثلة إلى حدٍ ما، مع احتمال منخفض لاختيار

المســتجيبين للخيارات الواقعة عند أطراف الســمة واحتمال عالٍ عند نقطة ما على النطاق المناســب لخيار الاستجابة للبند المعني. بالنسبة لخيارات الاستجابة الأعلى، يجب أن تظهر القمة عند مستويات أعلى من نطاق السمة مقارنةً بخيارات الاستجابة الأقل.

يمكننــا تصويــر احتمالية اختيــار كل خيار من خيارات الاســتجابة، بناءً على مســتوى الســمة التي يتم قياسها لدى المستجيب، من خلال سلســلة من منحنيات الاستجابة للفئة category-response curves. قــد تبدو مجموعة كاملة من منحنيات الاســتجابة للفئة الافتراضية لخمسة خيارات استجابة مستخدمة في المثال السابق بشكل يشبه ما هو موضح في الشكل رقم (٧-٧).

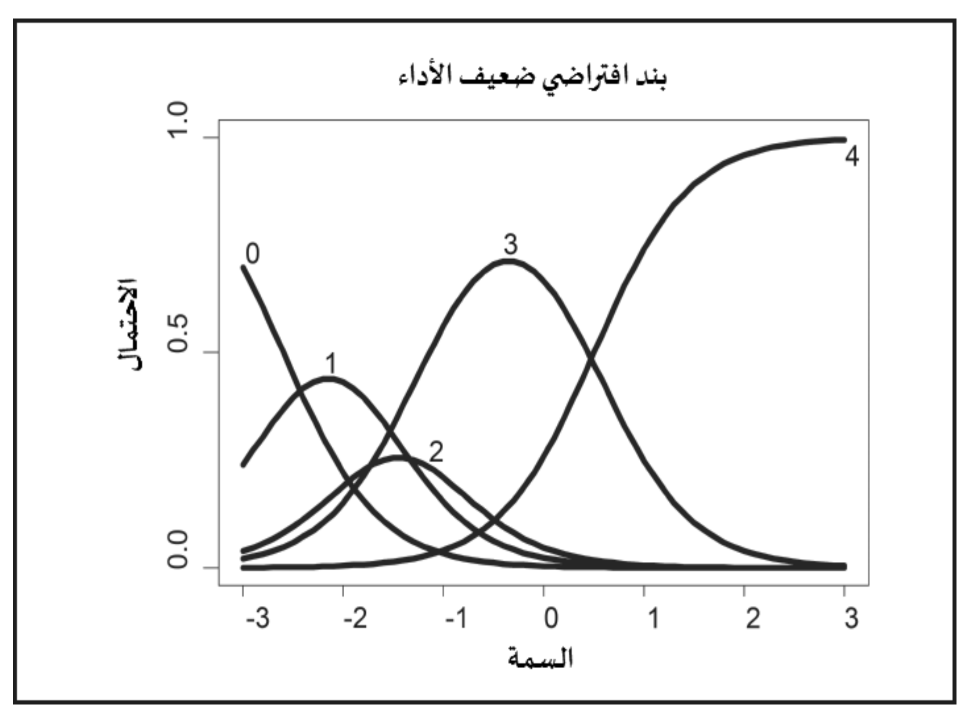
في الرسم التوضيحي المثالي الموضح في الشكل رقم (٧-٧)، يتوافق كل منحنى مُعلَّم برقم مع خيار الاستجابة في مثال الدوخة الصباحية الموضح سابقًا. يمثل خط الأساس قوة السمة التي يتم قياســها من خلال البند. ويتركز مقياس القدرة عند الصفر ويمكن تفسيره بطريقة تشبه الدرجات المعيارية.



شكل (٧-٧): منحنيات استجابة الفئة المثالية لخمسة خيارات استجابة

لاحظ أن هذا الشـــكل التوضيحي هو مثال مثالي وليس شـــكلاً بيانيًا حقيقيًا مبني على بيانات فعلية. يحتفظ هذا المثال بخاصية منحنيات استجابة البند الحقيقية: بالنسبة لأي مســـتوى من الســـمة، فإن مجموع المنحنيات المختلفة عند تلك النقطة على خط الأســـاس يســـاوي واحد. لذلك، لأي خط رأسي مرســـوم عند أي نقطة على مقياس السمة، فإن القيم (من المحور الواقع على الجهة اليسرى) للنقاط التي يعبر بها ذلك الخط منحنيات استجابة الفئة ســـيكون مجموعها يساوي الواحد الصحيح. بمعنى آخر، فإن احتمال اختيار استجابة ما يساوي الواحد الصحيح لأي مجيب بمستوى من السمة محل القياس.

يوضح الشكل رقم (٧-٨) مثالاً أقل مثالية ولكنه لا يزال افتراضيًا. يشير موقع منحنى كل خيـــار من خيارات الاســـتجابة الواقعة عـــلى امتداد بُعد الســـمة إلى "صعوبة" ذلك الخيـــار. ويمكننا أن نـــرى أن القمة لكل منحنى من منحنيات خيار الاســـتجابة المتتالية مرتبة كما كنا نتوقع، وتقع خيارات الاستجابة التي تشير إلى دوخة صباحية أقل على يسار خيارات الاستجابة التي تشير لدوخة صباحية أكثر. وهذا يعني أن الإشارة إلى المزيد من الدوار مرتبط بمنحنى يقع إلى الأعلى (أي إلى أقصى اليمين) على امتداد نطاق السمة، كما كنا نأمل. خاصية أخرى ملحوظة في الشكل رقم (٧-٨) هي أن منحنيات الاستجابة تبدو مزدحمة نحو الجانب الأيسر من المقياس. وهذا يعني أن الســـؤال يميز بشـــكل أكثر دقة بين الأشـــخاص الذين يعانون من دوخة قليلة نسبيًا مقارنةً بالأشخاص الذين يعانون من الدوار كثيرًا. إذا أخذنا بعين الاعتبار النطاق على مقياس السمة الواقع بين (-٠,٥ و-١,٥)، فإننا نرى أن جميع خيارات الاســـتجابة الخمسة يتم اختيارها أحياناً من قبل المستجيبين الذيـــن لديهم ذلك النطاق من قيم الدوخة، على الرغم من أن خيارات الاســـتجابة ١ و٣ يتم اختيارها في أغلب الأحيان. علاوة على ذلك، من المحتمل أن يختار من هم في النهاية العليا (أقصى اليمين) الخيار ٣ ("معظم الوقت")، في حين أن الموجودين في الطرف السفلي (أقصى اليســـار) سيختارون الخيار ١ ("نادرًا"). وبالتالي، حتى ضمن هذا النطاق الضيق، لا يزال بإمكاننا إجراء تمايز بين الأشخاص الذين هم أعلى وأقل فيما يتعلق بالدوخة.



شكل (٧-٨): مثال افتراضي لمنحنيات استجابة الفئة مزدحمة عند الطرف السفلي (الجهة اليسرى) من مقياس السمة

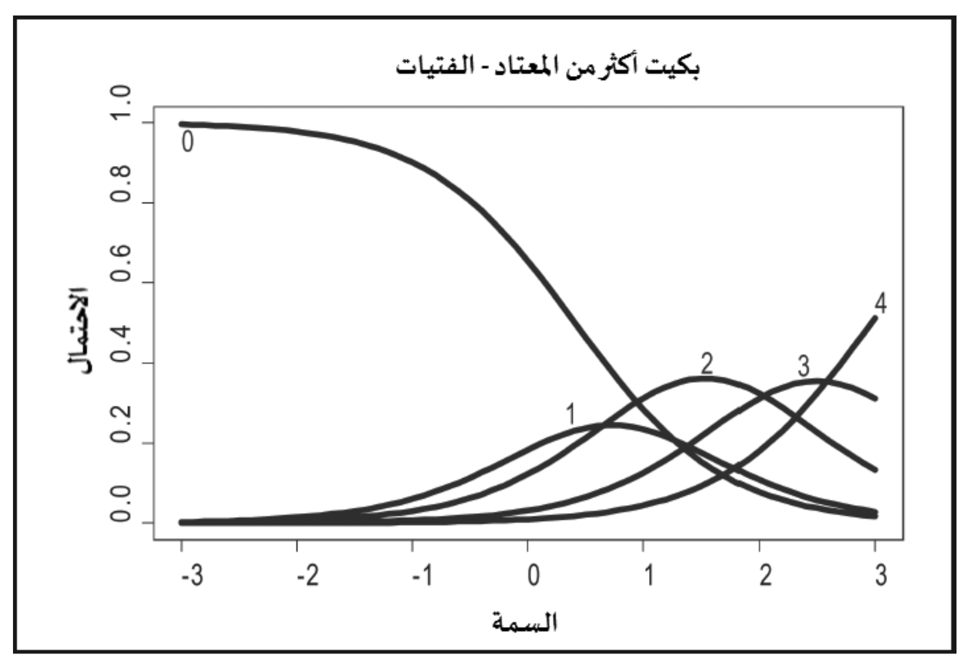
في المقابــل، إذا نظرنــا إلى موقع أبعد على مقياس الســمة، في حدود مــا بين ٢,٠ و٣,٠، فإن الاحتمال الغالب هو أن المســتجيبين الذين لديهم مستويات دوار في أي موقع بين تلك القيمتين ســيختارون الخيــار ٤ («تقريباً دائمًا»). في الواقع، «تقريباً دائمًا» هي الاســتجابة الأكثر احتمالًا لأي مســتجيب يحقق أعلى من ٠,٨ على مقياس الســمة. وبالتالي، يوفر هذا البند تمييزًا بسيطًا بين الأشخاص الذين يحققون ما بين ٠,٨ و٣,٠ بالنسبة لمستويات الدوخة لديهــم. منحنيات الاســتجابة للبند في الشــكل (٧-٨) لديها الكثير لتخــبر عنه. لاحظ أن منحنى اســتجابة الخيار ٢ («بعض الوقت») تتداخل قليلاً مع منحنيي الاســتجابة للخيار ١ («نــادرًا») والخيار ٣ («معظم الوقت»). لا يوجد فعليًا جزء من نطاق القدرة الذي يكون فيه خيار الاســتجابة ٢«بعض الوقت» هو الأكثر احتمالًا للاختيار. وبشــكل جوهري، فإن خيار الاســتجابة ٢ «بعض الوقــت» لا يقوم بأي شيء لا يمكن لخياري الاســتجابة الواقعين

على جانبيه أن ينجزاه. لا يزال من المرجح أن يقوم المســتجيب الذي حصل على درجة من الدوخــة الصباحية تعادل الدرجة (-١,٥) على مقياس الســمة (النقطة التي من المرجح أن يختار المجيبون عليها خيار الاستجابة ٢ باختيار الخيار ١ أو ٣. إن اختيار الخيار ٢ في هذه الحالة الافتراضية يؤدي إلى الغموض بين الدوخة النادرة والدوخة في بعض الأحيان. بالأخذ في الاعتبار هذا البند بشــكل منعزل، وبغض النظر عن أي بنود أخرى قد يتم تضمينها مع هذا البند في المقياس، فإن إســقاط خيار الاستجابة ٢ «بعض الوقت» من المحتمل أن يكون له معنى.

شــكل المنحنيــات لخيــارات الاســتجابة يعطينــا معلومــات حــول خاصيــة التمييز discrimination. وهذا يبدو أكثر وضوحاً بالنســبة للمنحنى الأخير، والذي يشبه إلى حد كبير المنحنيات التي قمنا بفحصها المتعلقة بنموذج الاســتجابة الثنائية (اجتياز/فشــل) ذي الثلاث معلمات (كما هو موضح في الشــكل (٧-٣)). المنحنى الموجود في أقصى اليســار هو نوع من صورة معكوســة طبق الأصــل لما رأيناه في نموذج المعلــمات الثلاث من حيث إنه يصل إلى أعلى مستوياته في أقصى اليسار وينحدر مع التقدم للأعلى في مقياس السمة، ولكن مرة أخرى، فإن انحدار ميل المنحنى يدل على درجة تمييزه للســمة. بالنســبة للمنحنيات الموجودة بين المنحنيات الطرفية، يشير الانحدار (أو ربما بدقة أكثر، التفرطح kurtosis) إلى درجة التمييز. خيارات الاســتجابة التي تكون منحنياتها مرتفعة ومدببة تميز بشكل أفضل من تلك التي تكون منحنياتها أكثر اتساعًا وانبساطاً. في الشكل رقم (٧-٨)، خيار الاستجابة ٢ ("بعــض الوقــت") له منحنى على شــكل القرفصــاء إلى حد ما، مما يشــير إلى أن خيار الاستجابة لا يميز بشكل جيد. هذا يدعم كذلك فكرة أن حذف هذا الخيار قد يكون مناسباً.

لاختتام هذا القســم، ســننظر في منحنيات الاســتجابة للفئات (الخيارات) الناتجة عن بيانات فعلية (تم توفيرها واستخدامها بإذن من زملائي دارين دي ولت Darren DeWalt وديفيد ثيســن David Thissen من جامعة UNC) وســوف نســتخدم هــذه المنحنيات لتوضيح الغرض الآخر الذي يمكنها تقديمه. يمثل الشــكل رقــم (٧-٩) بنداً تم تطبيقه على مجموعــة من المــرضى الأطفال من الإناث كجــزء من جهد لتطوير مجموعــة بنود لتقييم الاكتئــاب. هذا البند يطلب من الأطفال الإشــارة إلى عدد المرات التــي بكوا فيها أكثر من المعتــاد في المــاضي القريب. على نحو مشــابه لكن غــير مطابق لأمثلتنا الســابقة، كانت

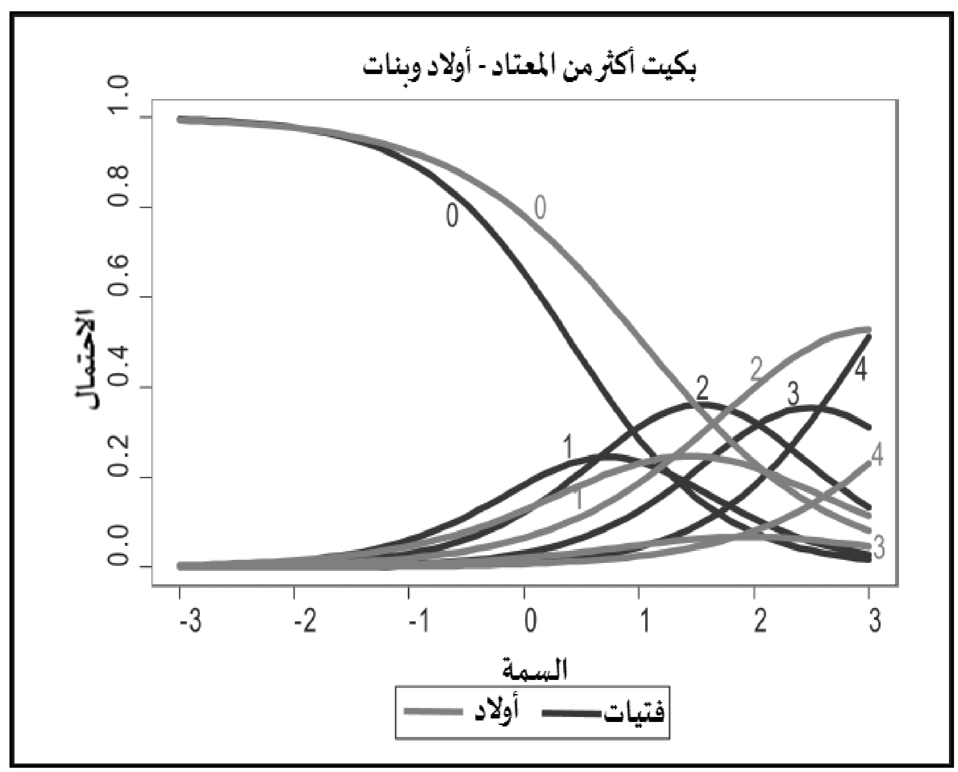
خيارات الاســتجابة هي صفر ("أبدًا") و١ ("أبدًا تقريبًــا") و٢ ("أحيانًا") و٣ ("غالباً") و٤ ("دائمًا تقريبًا"). يشــير تجمع المنحنيات إلى اليمين إلى أن المشــاركين يجب أن يكونوا على درجة عالية من السمة (أي الاكتئاب) من أجل تقديم إجابة بخلاف "أبدًا".



شكل (٧-٩): منحنيات استجابة الفئة للفتيات للبند المتعلق بالبكاء أكثر من المعتاد

يبيّن الشــكل رقم (٧-١٠) استجابات الأولاد واستجابات الفتيات معاً على نفس الشكل. على الرغم من أن هذا الشكل مزدحم للغاية، إلا أنه يكشف أن منحنيات الأولاد، والفتيات متشــابهة عمومًا ولكنها ليست متطابقة بأي حال من الأحوال. ولجعل الفارق الحاسم بين اســتجابات الأولاد والفتيات أكثر وضوحًا، في الشكل رقم (٧-١١)، تم تحرير المنحنيات التي تمثل خيارات الاســتجابات الوســطى الثلاثة، وتم الإبقاء فقط على تلك المنحنيات الخاصة بـــخيارات الاســتجابة صفر "أبدًا" و٤ "دائمًا تقريبًا". بالنســبة للخيار صفر، يكون منحنى الأولاد أعلى من الفتيات، أما بالنسبة للخيار ٤، فإن الفتيات "أعلى من الأولاد".

كيــف يجب أن نفــسر هذا الاختلاف؟ هذا الاختلاف يشــير إلى أنه خلال المدى الكامل لقيــم الســمة (الاكتئاب)، فإن الأولاد أكــثر أرجحية من الفتيات للقــول إنهم لم يبكوا (أي اختيار خيار الاستجابة "أبداً") أكثر من المعتاد. في الطرف العلوي من مقياس السمة، تكون الفتيات أكثر احتمالية من الأولاد للاعتراف بأنهم يبكون أكثر من المعتاد "دائمًا تقريبًا" خلال الفترة الزمنية المعنية.



شكل (٧-١٠): منحنيات استجابة للفئة متداخلة للفتيان والفتيات للبند المتعلق بالبكاء أكثر من المعتاد

يشير هذا النمط إلى أحد أمرين: إما (١) أن تكون الفتيات أكثر اكتئابًا من الأولاد أو (٢) يولّد مســتوى منخفض من الاكتئاب استجابات أعلى لدى الفتيات منها لدى الأولاد. أي إن البند يعكس الفروق الحقيقية بين الجنسين بين الأولاد والفتيات أو أن البند لا يعمل بشكل متكافئ بين الجنسين. الشك لدينا هو أن الإقرار بالبكاء قد يكون أكثر صعوبة لدى الفتيان

منــه لــدى الفتيات. ونتيجة لذلك، فــإن الأولاد الذين يختارون أي خيار اســتجابة بخلاف «أبداً» قد يكونون بالفعل أكثر اكتئابًا من الفتيات اللائي يخترن خيار الاستجابة نفسه. وهذا يشــير إلى أن البند يظهــر أداء البند التفاضــلي differential item functioning (DIF). بمعنى أن المجموعات المختلفة من المجيبين (الأولاد، والفتيات في هذه الحالة) الذين لديهم مســتويات مماثلة من الســمة يحصلون على درجات مختلفة للبند، مما يشير إلى أن البند يحدد على مقياس الســمة بشــكل مختلف كدالة في مجموعة الجنس. من خلال النظر إلى البنود الأخرى التي تقيس الاكتئاب والتي لا تســأل عــن البكاء ومن خلال ملاحظة النتائج المتشــابهة بشكل أســاسي للبنين والبنات، وجدنا دليلًا إضافيًا على أداء البند التفاضلي DIF للبنــد المعني. بينما توجد طرق أخــرى لتحديد أداء البند التفاضلي DIF، يجد العديد من الباحثــين المعلومات المرئية التي تنتجها حزم نظرية الاســتجابة للبنــد IRT لتكون مقنعة بشكل خاص.

خلاصة القول، يمكن أن يوفر فحص منحنيات استجابة الفئات معلومات من عدة أنواع حول كيفية أداء بند ما. ومن بين الأشياء الكثيرة التي يمكن أن يكشف عنها فحص منحنيات الاستجابة للفئة (أ) معلومات حول مدى التمثيل الكلي لبند ما للمجموعة الكاملة للخاصية التي يتم قياسها، (ب) ما إذا كانت خيارات الاستجابة المرتبة تمثل بدقة المناطق الأعلى من مقياس السمة بشكل تدريجي، (ج) ما إذا كانت جميع خيارات الاستجابة ضرورية ومفيدة، (د) مدى قدرة البند على التمييز بشــكل جيد عند مستويات مختلفة من السمة، و(هـ) ما إذا كان البند يظهر أداء البند التفاضلي DIF. يمكن الحصول على أنواع أخرى من المخرجات الرســومية من العديد من حزم نظرية الاســتجابة للبند، ويكون التمثيل المرئي لأداء العنصر من بين أكثر ميزات IRT فائدة.

ثيتــا والاختبار التكيفــي المحوســب (Theta and Computerized Adaptive Testing (CAT:

في قلــب كل أســاليب القياس، هناك مفهوم أن قوة الســمة التي يتم قياســها تختلف باختــلاف الأفــراد أو الوقــت، ونظرية الاســتجابة للبند IRT ليســت اســتثناءً من ذلك.

وفي وقت ســابق من هذا الفصل، تم استخدام مصطلح "القدرة" لوصف مثل هذه السمة. وغالبًا ما يشــار إلى هذه الســمة، بصرف النظر عن محتواها الموضوعي، بـ "ثيتا" في سياق نظرية الاســتجابة للبند IRT ويمكن الإشــارة إليهــا بالحرف اليوناني الكبــير Θ. كما ذكر ســابقًا في هذا الفصل، يمكن إنشاء البنود بحيث يكون لها مستويات مختلفة من الصعوبة. وبشكل أساسي، تتم معايرة الصعوبة من حيث ثيتا. وبالتالي، فإن الاستجابة الإيجابية لبند أكثر صعوبة (أو اختيار خيار استجابة يمثل قيمة أعلى) تعني أن المستجيب لديه قيمة ثيتا أعلى (أي أنه يمتلك مســتوى أعلى من الســمة التي يقيمها البند) نسبيًا للمستجيبين الذين يؤيــدون فقط البنــود المرتبطة بقيم ثيتا الأقل. بالنســبة لكل بند، يمكــن لمحلل البيانات تحديد مستوى (أو نطاق) قيم ثيتا التي يميز البند عندها بين المشاركين بمستويات مختلفة من السمة. بعض البنود قد تعمل بشكل جيد عبر مجموعة واسعة من قيم ثيتا، بينما بنود أخرى ستعمل بشكل جيد فقط عبر نطاق ضيق من ثيتا.

إن بنداً افتراضياً مثل «أنا قلق على مســتقبلي» مقروناً بخيارات الاســتجابة (٠) «أبدًا» (١) «القليــل مــن الوقت» (٢) «معظم الوقت» و(٣)) «دائمًا»، قــد يكون قادرًا على توفير لمحة عامة للسمة التي يقيمها عبر نطاق واسع من ثيتا. ومع ذلك، قد لا يفرق بشكل جيد للغاية بين التقديرات الأصغر للخاصية، مثل نادراً rarely ما تشعر بالمحتوى وقلّما seldom (بمعنــى أكــثر بقليل من نادراً) تشــعر بالمحتــوى. قد يختار الأشــخاص في كل من هذين المســتويين الأخيرين من القلق على المســتقبل "القليل من الوقت"، وبالتالي لا يمكن للبند تمييز الفرق الدقيق في "القلق على المستقبل" بينهما. ولكن بالنسبة لبعض التطبيقات، قد يكــون التمييز بين الأفراد أصحاب الخيارات الدقيقة "نــادراً ما" rarely و"قلما" seldom يعتبر أمرا مهمًا. وبالطبع، يمكن إنشاء بنود إضافية تركز بمزيد من التفصيل على جزء أقصر من النطاق الكامل للســمة. قد يكون لهذا البند خيارات إجابة مثل "أبدًا"، و"نادرًا جدًا"، و"نادرًا"، و" قلما"، و"من حين لآخر". أو قد تتم إعادة صياغة أصل البند نفسه لاستهداف مســتوى مختلف من ثيتا. على ســبيل المثال، "لدي شــعور بالفزع حيال مســتقبلي" يثير بوضوح المخاطر. أيٌّ من هذين النموذجين البديلين للبند قد يركّز على المســتويات الأعلى لسمة القلق أكثر من البند الأصلي.

في الواقــع، يمكن تطوير مجموعــة من البنود من هذا النوع تســتهدف أجزاء مختلفة من سلســلة قيم ثيتا، بعضها يركز في المقام الأول على المستويات المنخفضة، والبعض الآخر على المســتويات المتوســطة، والبعض الآخر على مســتويات عالية من القلق. من المحتمل أن تتفوق مجموعة كبيرة ومصاغة بشــكل متقن من هذه البنود على عدد أقل من البنود العامة. وهذا يعني أن هذه المجموعة الموسعة من البنود يمكن أن تحل الاختلافات الأصغر في ثيتا بدقة أكبر (أي تفريق المجيبين الذين كانت مســتويات قلقهم أكثر تشابهًا) بالإضافة إلى تمديــد النطاق الكامل لثيتا (أي، تضمين بنود مســتهدفة لعدة مســتويات مختلفة من القلــق تتراوح من منخفض للغاية إلى مرتفع للغاية). المشــكلة في هذه الإســتراتيجية هي أنهــا تتطلب عددًا كبــيرًا من البنود. علاوة على ذلك، قد يضطر أي مجيب معين إلى إكمال عدد كبير من البنود التي كانت خارج نطاق ثيتا الخاص به (على ســبيل المثال، يتعلق ذلك بمســتويات أعلى أو أقل من الســمة من تلك التي يمتلكها المجيب). هذا من المحتمل أن يكون له ضريبة على المجيبين الذين قد يشــعرون بالضجر من اســتكمال العديد من البنود والذيــن ربمــا يكونون منزعجين من البنــود التي لا تنطبق على ظروفهــم. إن إدراك هذه المشاكل المحتملة قد يلزم مطوري المقياس إما باختيار مجموعة أقصر من البنود التي قد لا تؤدي مهمة جيدة لاكتشاف الاختلافات الأكثر دقة في ثيتا، أو باختيار أداة أطول تميز بدقة أكبر ولكن من المحتمل أن تكون أكثر كلفة.

يقدم الاختبار التكيفي المحوســب (CAT) علاجًا ممكنًــا لهذه المعضلة. في CAT، تم تطويــر بنك كبير مــن البنود، وتتم معايرة كل منها بعناية بحيــث تكون قيمة theta التي تتوافق معها معروفة. يكمل المجيبون البنود عبر الكمبيوتر، مع تحديد خوارزمية مســبقة الضبط بشكل تكيفي لتحديد عدد البنود المعروضة وترتيبها. البنود القليلة الأولى عادةً ما تكون عامة إلى حد ما، ولا تسفر إلا عن فكرة تقريبية عن مستوى المجيب من السمة التي يجــري تقييمها. اســتنادًا إلى الإجابات على تلك البنودالأولية، يمكــن لخوارزمية الكمبيوتر تقديــم البنود التالية بشــكل انتقائي، واختيار تلك البنود الموجــودة ضمن مجموعة البنود المتوفرة الأكثر صلة والغنية بالمعلومات لمستوى ثيتا التي يبديها المستجيب المحدد. بمعنى، يمكن للكمبيوتر تقديم البنود التي تغطي منطقة من السلسلة المتصلة لقيم ثيتا القريبة من مســتوى السمة التي كشفت عنها الاستجابات السابقة. لذلك، على سبيل المثال، إذا كانت

الاســتجابات الأولية تشــير إلى أن المجيب يقلق بشأن مســتقبله أكثر من معظم الأشخاص، فيمكن اختيار البنود التي تركز على مستويات أعلى من القلق بينما يمكن حذف البنود التي تســتهدف مستويات منخفضة للغاية. مع استمرار التقييم، يمكن للكمبيوتر تحسين اختيار البنود، مع التأكد من أن البنود المعروضة ذات صلة بمستوى سمة المجيب وكافية في العدد لتوفير وصف دقيق.

بالمعنــى الحــرفي للكلمة، لا يعتمــد الاختبار التكيفــي من هذا النوع على اســتخدام نهــج نظرية الاســتجابة للبند IRT. في الواقــع، قبل تطوير هذه الأســاليب بفترة طويلة، اســتخدمت اختبارات الــذكاء الموحدة اســتراتيجية مماثلة لطرح بعض الأســئلة "محددة الموقع" (questions «locator») لتحديد أي من الأسئلة المتعلقة بمنطقة معينة من الذكاء يجب طرحها وتحديد عددها. ومع ذلك، فإن القدرة التي يوفرها نهج IRT لمعايرة البنود لمســتويات محددة من ثيتا (وللتعبير عن هــذه المعلومات من الناحية الكمية)، إلى جانب قدرة أجهزة الكمبيوتر على الاســتفادة في الوقت الفعلي من مجموعة الاستجابات المقدمة واستخدام تلك البيانات لتحديد البند التالي الذي سيتم تقديمه، زادت بشكل كبير من فائدة واعتماد الاختبار التكيفي.

الاختبار التكيفي المحوسب CAT يمكن أن يكون جذاباً للغاية عندما يكون ذلك ممكنًا. وهو يتطلب وجود بنك كبير من البنود التي تم اختبارها بشكل كافٍ لتوفير معايرة جديرة بالثقة. أيضًا، تحتاج خوارزميات CAT إلى حساب خصائص البند بخلاف نتائج المعايرة. على ســبيل المثال، قد يؤدي الاختيار على أســاس ثيتا إلى التضحية دون قصد بنطاق المحتوى إذا تم استبعاد البنود التي تمثل بنوداً مهمة للمفهوم البنائي (construct) قيد الفحص. لا تزال الدراســات التجريبية التي تقارن CAT مع استخدام مجموعات البنود الثابتة نادرة نسبيًا، على الرغم من أن التقارير المتاحة تشــير إلى وجود ميزة لـ CAT مقارنةً بمجموعات البنود الثابتة ذات الحجــم المتواضع (مثــل Choi, Reise, Pilkonis, Hays, & Cella, 2010). يبدو من المحتمل أنه كلما كان نطاق المحتوى موضع اهتمام أكبر، وكلما كانت مســتويات الســمة أكثر تقلبًا بين شرائح مختلفة من المجتمع، فإنه من الممكن أن تكون فوائد الاختبار التكيفي المحوسب CAT أكبر.

للحصــول على نظرة أكثر شــمولاً على الاختبار التكيفي المحوســب CAT، يمكن للقراء الرجــوع إلى (Bjorner, Chang, Thissen, and Reeve (2007. المجلد الأخير الذي يمكن للقراء اســتخدامه لتعزيز فهمهم لمجموعة واســعة من موضوعات نهج نظرية الاســتجابة للبنــد IRT هو كتيب نمذجة نظرية الاســتجابة للبند: تطبيقــات لتقييم الأداء النموذجي Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment (Reise & Revicki, 2015).

تعقيدات نظرية الاستجابة للبند IRT:

على الرغم من أنها جذابة للغاية، فإن نظرية الاســتجابة للبند IRT ليســت حلاً سريعًا لمشاكل القياس. مثل نظرية القياس الكلاسيكية CTT، لا تحدد IRT خصائص البنود؛ إنها مجــرد قياس كمي للبنــود. لذا، فإن التكنولوجيا في حد ذاتها تســمح للباحث بتقييم أداء البند، لكنها لا تتســبب بشــكل مباشر في كتابة بنود أفضل أو جعل البنود التي تم إنشاؤها بشــكل سيء تعمل فجأة بشــكل جيــد. أيضاً، يمكــن أن تكون عملية التقييم شــاقة عند استخدام طرق تعتمد على IRT. القياس الكلاسيكي يضحي بالدقة لأجل البساطة من خلال تبنــي تصورات أقل تباينًــا ولكن أكثر تعقيدًا لمصادر الخطأ. بينما تقوم نظرية الاســتجابة للبند IRT بالخيار العكسي، واكتســاب الدقة مقابل التضحية بالبســاطة. علاوة على ذلك، يتطلــب تطبيــق طرق IRT درجة كبــيرة من قرار الخبراء. لا تزال هــذه الطرق في مرحلة التطوير النشطة، مع ظهور مشكلات جديدة وتقديم حلول جديدة.

وللحصول على الثقة بأن خصائص البنود يتم تقييمها بشكل مستقل عن خصائص العينة التي تتم دراستها، وهو الهدف الأساسي لـ IRT، يجب على المرء أن يثبت أن هذه الخصائص متســقة عبر مجموعة واســعة من المشــاركين الذين يختلفون في نواحٍ مختلفة، بما في ذلك مستوى القدرة. ومن المهم ألا ترتبط خصائص البند بخصائص العينة الأخرى المستقلة عن السمة، مثل الجنس أو العمر أو المتغيرات الأخرى التي يجب أن تكون غير مرتبطة مع تلك التي يتم قياسها. يجب أن تختلف درجات البنود فقط عندما تتغير السمة محل الاهتمام، وليس بســبب الاختلافات في أي متغيرات خارجية المنشــأ extraneous variables. لذلك، على ســبيل المثال، إذا افترضنا أن القدرة الإملائية spelling لا علاقة لها بالجنس، فســيتعين

علينــا إثبات أن الأولاد والبنات ذوي القدرة المتســاوية ســيكون لديهم نفس الاحتمال في اجتيــاز بنــد ما. إذا لم يكن هذا صحيحًا، فســيؤثر نوع الجنس أو بعــض العوامل الأخرى بخلاف القدرة الإملائية على البند. أيضًا، كما هو الحال مع النظرية الكلاســيكية، يجب أن تشــترك البنود التي يتــم فحصها في مجموعة (أي تكوين مقيــاس لقياس المتغير) في متغير أساسي واحد فقط.

إحــدى نتائج هذه المتطلبات هي أنه يجب عــلى الباحثين الوصول إلى أعداد كبيرة من المجيبــين المتنوعين مــن أجل وصف البنود بدقة. الهدف الرئيــسي لـ IRT هو ربط البنود بمستويات محددة من الســمة التي يتم تقييمها. البنود الجيدة والعينات الكبيرة ضرورية لتحقيق ذلك. تتطلب الاستقلالية النظرية لخصائص البنود وخصائص العينة التي تعد سمة مميزة لـ IRT أن يكون قد تم تقييم البنود عبر النطاق الكامل للســمة محل الاهتمام وأن أداة القياس تتضمن بنوداً حساسة بشكل شامل لجميع مستويات تلك السمة.

كيف يتحقق مطور أدوات القياس من المستوى الحقيقي للسمة بطريقة تسمح بتكوين المنحنيات المميزة للبنود ICCs؟ بالعودة إلى تشــبيه المطرقة والجرس، كيف يمكنك تحديد القــوة من أجل تحديــد مقدار القوة التــي يتطلبها قرع الجرس لطــرق جزئية معينة من الجهــاز؟ في كثير من الحالات، إذا كان المســتوى الحقيقي للســمة معروفًــا بطريقة يمكن التحكــم فيهــا، فلن تكون هناك حاجة كبيرة لتطوير مقياس جديــد. من الناحية النظرية، نظرًا لاســتجابات عدد كبير من الأشخاص لمجموعة ثابتة من البنود، يجب أن يكون برنامج الكمبيوتر قادرًا على فرز الاختلافات الناشــئة عن خصائص البند مقابل خصائص الشــخص. للعودة مرة أخرى إلى تشــبيه أجهزة الكرنفال (جهاز رنين الجرس وآلة الغمر)، إذا استخدم عدد كافٍ من الأشــخاص كلاً من الجهازين، على ســبيل المثال، يجب أن يكون من الممكن تحديــد أي من الجهازين أصعب، وأيضًــا، للحكم على مهارات الأفــراد في هاتين المهمتين. في الممارســة العملية، غالبًا ما تكــون هناك عملية تكرارية متداخلــة تتضمن إدارة البنود لقياس مستوى السمة لمجيبين معينين، ثم استخدام تقدير السمة كدليل في تحديد خصائص البنود الأخرى. عندما يتم تحديد أفضل البنود على هذا الأساس، يمكن استخدامها للحصول على تقدير محسّن لمستويات الأفراد في السمة للجولة التالية من اختيار البنود، وهكذا.

نظرًا لطبيعة هذه العمليات، من السـهل أن نرى السـبب وراء تبنّي نظرية الاسـتجابة للبند IRT بحماس شـديد من قبل المؤسسات التجارية المعنية باختبار القدرات، مثل تلك التي تدير اختبارات سـجل الخريجين. يوفر التطبيق والتقييم المستمر للبنود بمرور الوقت أساسًـا ممتازًا للعثور على البنود التي تتسم خصائصها بالاستقرار خلال التباين في مجموعة واسعة من خصائص المستجيبين الأخرى.

أحد التعقيدات الأخرى لنظرية الاسـتجابة للبند IRT يتعلق باسـتخدام بنود مختلفة لقياس الشيء نفسـه في أوقات مختلفة. في وقت سـابق، ذكرت باختصار الاختبار التكيفي المحوسـب (CAT) (Van der Linden & Glas, 2000) كنهـج لجمـع البيانات المرتبطة غالبًا بـ IRT. في CAT، تتم معايرة البنود بشـكل فردي لمقياس السـمة باسـتخدام طرق IRT، وبالتالي، يمكن اختيارها للتطبيق بحيث تتوافق مع مستوى السمة لمجيب معين. هذا يعني أن البنود المختلفة سـتكون مناسبة على النحو الأمثل لجوانب القوة المختلفة للسمة. يمكن أن تحدث مثل هذه الاختلافات إما بين الأشخاص الذين يختلفون فيما يتعلق بالسمة أو لفرد معين يتغير مستوى السمة الخاصة به.

وبالتالي، فإن استخدام الاختبار التكيفي المحوسب CAT غالباً ما يستلزم نتائج أو درجات مـن بنـود مختلفة تتـم مقارنتها مع بعضها البعض. على سـبيل المثـال، إذا تغيرت إحدى السـمات كنتيجة للتدخـل، فقد تختلف أفضل البنـود لقياس تلك السـمة في أي فرد بعد التدخـل عن ذي قبل. بالنسـبة للمسـتخدم لتقرير علمي يبني ادعـاء نجاح التدخل على أسـاس مقارنة بين البنود المختلفة عبر (بسبب الاختلافات الأولية في السمة) وداخل (بسبب التغيـرات الناتجة عـن التدخل) الأفراد، فإن الوثوق في البيانات قـد يتطلب الإيمان بذلك. يتمتع معظمنا بخبرة أكبر بكثير (وربما بالراحة) في الظروف التي تم فيها إعطاء البنود نفسها قبـل وبعد التدخل مقارنةً بالسـيناريو الذي تم وصفه للتـو. عندما يتم تطبيق طرق IRT بشكل صحيح وتتم معايرة العناصر بعناية بالنسبة لسلسلة قيم السمة على أساس البيانات التـي تم جمعها من عينـات كبيرة ومتنوعة، فإنه يمكن تحويـل الدرجات الناتجة من بنود مختلفة إلى مقياس سمة شائع بنفس الطريقة التي يمكن من خلالها تحويل قياسات بالبوصة أو بالمليمتر إلى مقياس شـائع للطول. عندما تكـون الدقة التي يمكن بها تعيين البنود على بُعد السمة منخفضة، فإن تفسير النتائج بناءً على مجموعات بنود مختلفة قد يكون مشكلة.

علاوة على ذلك، نظرًا لأن مستخدمي تقارير الأبحاث قد لا يفهمون تمامًا جميع المشكلات التــي ينطوي عليها نوع مقارنــات البنود المختلفة التي قد ينطــوي عليها الاختبار التكيفي المحوسب CAT، فقد يواجه مؤلفو هذه التقارير درجة أكبر من الشك.

الخلاصة:

إن أســاليب القياس القائمة على نظرية الاســتجابة للبند IRT لديها العديد من الميزات الجذابة. ولكن إنشــاء البنود الجيدة هو عمل شــاق، بغض النظر عن الإطار النظري الذي يوجه العملية. إن كتابة البنود التي تقيس الســمة قيد الدراســة باتســاق والتي لا تراعي خصائص المستجيب الأخرى ليست مهمة سهلة. بينما في المقاييس التي تعتمد على النظرية الكلاسيكية، فإن وجود بنود متعددة يمكن أن يعوض عيوب بعض هذه البنود إلى حدٍ ما، في حين أن المنطق وراء نظرية الاســتجابة للبند أن كل بند يتم الحكم عليه على حده. بسبب أنه يمكن للمرء أن يكتشف، من خلال فحص المنحنيات المميزة للبنود (ICCs)، على سبيل المثــال، كــون أداء أحد البنود جيدًا لا يعني أن ذلك المرء ســيفعل ذلك. إن امتلاك معرفة مستقلة موثوقة بالخصائص التي يتم قياسها هو أحد متطلبات نظرية الاستجابة للبند التي يصعــب الوفاء بها بدقة، ولكن يمكن تقريبها بشــكل كافٍ بالاختبــارات المتكررة للعينات الكبيرة وغير المتجانســة. وبالتالي عندما لا يكون هذا خيارًا، فقد يكون من الصعب للغاية إقناع النقاد بأن الافتراض قد تحقق بشكل كاف.

وجهة نظري الشخصية هي أنه عندما تكون افتراضات نظرية القياس الكلاسيكية CTT قابلة للتطبيق (أي، حيث يتم تصميم البنود كمؤشرات مكافئة لمتغير أســاسي مشترك)، فإن قابليــة التتبــع والأداء لهذه المقاييــس تجعلها خيارات جذابة. مــن ناحية أخرى، في حالة توفــر عينات كبيرة (والتــي يُعد عندها الاختبار التكيفي المحوســب CAT طريقة مفضلة للتطبيق)، أو في حالة أن الدرجات المكافئة في دراســات متعددة تكون ذات أهمية قصوى، فإن الأساليب القائمة على IRT في ظل وجود التعقيد المضاف قد تكون هي الخيار الأفضل. ومع ذلك، فإن مجرد اســتخدام هذه الأساليب ليس ضمانًا للمنتج النهائي المنشود. ويجب على الباحث أن يثبت أن افتراضات الطريقة المختارة قد تم الوفاء بها ضمن حدود مقبولة وأن ثبات reliability أداة القياس الناتجة وصدقها validity قابلة للتحقق منها تجريبياً.

هل تجعل نظرية الاستجابة للبند الأساليب الكلاسيكية قديمة؟ يدرك العديد من مؤيدي نظرية الاســتجابة للبند أن لكل من نظرية القياس الكلاسيكية ونظرية الاستجابة للبند دور تلعبه. على ســبيل المثال، ذكر Embretson & Hershberger (1999)، في أول توصياتهما المتعلقــة بالتغييرات في أســاليب القياس الحاليــة، أنه يجب دمج نظرية الاســتجابة للبند والأساليب الكلاسيكية CTT في نهج شامل لقضايا القياس. (ص ٢٥٢).

على الأقل، خلصت دراستان تجريبيتان كبيرتان (Fan, 1998; Stage, 2003) التي قارنت خصائــص الاختبارات التربوية القائمة على الأســاليب الكلاســيكية ونظرية الاســتجابة للبند إلى أن الأســاليب الكلاسيكية أدت بشــكل جيد أو أفضل من الأســاليب القائمة على نظرية الاســتجابة للبند. استندت دراسة Fan (1998) إلى ٤٠ عينة كل منها مكونة من ١٠٠٠ فرد، مأخــوذة مــن مجموعة تضم حوالي ١٩٣٠٠٠تلميذ من تكســاس تــم اختبارهم في مهارات القراءة والرياضيات. وقد أظهرت خصائص البند والشخص اتساقًا كبيرًا عبر الأساليب. قامت Stage (2003) بمقارنة الطرق التقليدية وطرق نظرية الاستجابة للبند باستخدام بيانات من اختبارات القدرات القياسية السويدية Swedish Standard Aptitude Tests. واستناداً إلى عينة من ٢٤٤١ من المتقدمين للاختبارات تم اختيارهم عشــوائياً من مجموعة من ٨٢٥٠٦، خلصــت إلى أنــه في حين أن نموذج نظرية الاســتجابة للبند المكون مــن ثلاثة معلمات كان متوافقاً (fit)مع البيانات بشــكل ضعيف، فإن نموذجًا قائماً على نظرية الاختبار الكلاســيكية كان جيدًا. في الآونة الأخيرة، أجرى Silvestro-Tipay (2009) دراســة مماثلة ولكنها أصغر باستخدام درجات اختبار لـ ٣٢٦ طالباً جامعياً من طلبة السنة الأولى. وقد خلص إلى أن:

> النتائج هنا توضح ببســاطة أن إطاري القياس ينتجــان إحصائيات متماثلة لكل بند وشــخص سواء من حيث قابلية المقارنة بين إحصاءات البند والشخص، ومستوى صعوبة البنود، والاتساق الداخلي، وأداء البند التمييزي بين إطاري العمل. (ص ٢٩).

وعــلى نفس المنوال، قدم باحثان مــن خدمة الاختبارات التربوية ومجلس الكلية مؤخرًا ورقة تصف نتائج دراسة محاكاة قارنت طرق نظرية الاستجابة للبند والأساليب الكلاسيكية مع تحديد نماذج الاختبار (Davey & Hendrickson, 2010). وخلصوا إلى أن "كلا النهجين يوفران مزيجًا مثيرًا للاهتمام من المزايا والعيوب النظرية والعملية، مع عدم ظهور أيٍ منهما كمفضل واضح على الورق". وقد لاحظوا كذلك أنه:

> كانت النتيجة الأكثر وضوحاً للدراسة هو الفرق البسيط نسبياً بين أداء طرق الكلاسيكية ونظرية الاستجابة للبند .... قد يعني التشابه في الأداء لتلك الأساليب أن اختيار الممارس يعتمد إلى حد كبير على التفضيل أو الملاءمة التشغيلية. (ص ٢)

لا ينبغي تفسير هذه التقارير على أنها تشير إلى أن أساليب نظرية الاستجابة للبند ليست لها مزايا في أي موقف. لا شيء من هذه الدراسات قطعي، وبعض هذه الاختلافات تم تحديدها عند مقارنة تلك الطرق. تحليلات نظرية الاستجابة للبند في بعض الحالات قدمت معلومات أكثر تفصيلاً من التحليلات الكلاسيكية. وكما رأينا في الرسوم البيانية للبنود متعددة الاستجابات، يمكن لطرق نظرية الاستجابة للبند أن تقدم رؤى مفيدة، مثل الإشارة إلى أن جميع خيارات الاستجابة ليست ضرورية أو أن البند لا يلتقط معلومات عبر النطاق الكامل للسمة التي يتم قياسها. توفر طرق نظرية الاستجابة للبند أيضًا ميزات نظرية واضحة (مثل استقلال خصائص البند عن خصائص العينة)، ومع تحسن الأساليب، قد يتم تقليص بعض العوائق العملية. ولكن كحد أدنى، تجادل هذه الدراسات أن طرق نظرية الاستجابة للبند ليست بالضرورة أفضل من طرق القياس الكلاسيكية. باختصار، إن تقارير زوال القياس الكلاسيكي سابقة لأوانها.

ولتناول مزايا وعيوب كلا الطريقتين، لاحظ Zickar and Broadfoot (2008):

> كما أن طرق نظرية القياس الكلاسيكية CTT لها قيودها الخاصة، لاحظ الباحثون وجود قيود شديدة على طرق نظرية الاستجابة للبند IRT التي تجعل استخدام أساليبها صعبة أو مستحيلة أو غير عملية في سيناريوهات معينة. تتضمن هذه القيود الحاجة إلى أحجام كبيرة للعينات، وافتراضات قوية لأحادية البعد unidimensionality [وهو ما ينطبق أيضًا على الأساليب الكلاسيكية]، وصعوبة تنفيذ البرامج. (ص ٤٨).

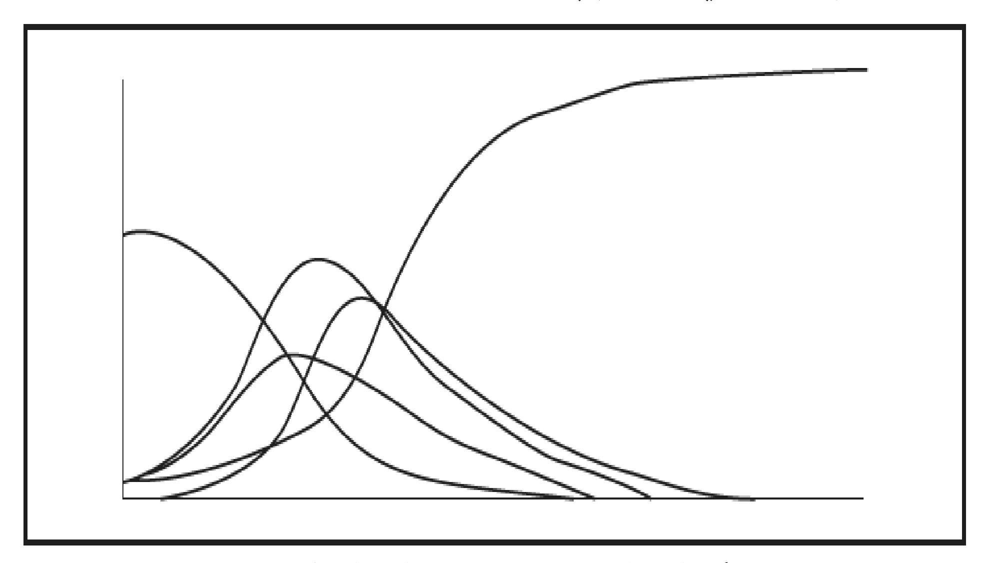
فيما بعد في نفس التقرير، يذكر هذان المؤلفان:

> على الرغم من أن الأســطورة الشــائعة أن نظرية القياس الكلاسيكية قد ماتــت، إلا أننا نعتقد أن هناك العديد من الســيناريوهات التي فيها يفضل اســتخدام نظرية القياس الكلاسيكية. معظم الأســباب التي يمكن تصنيفها تعود للقيود في البيانات التي قد تستبعد نظرية الاستجابة للبند والاعتبارات العملية التي قد تجعل نظرية القياس الكلاسيكية أكثر تفضيلًا. (ص ٥٠).

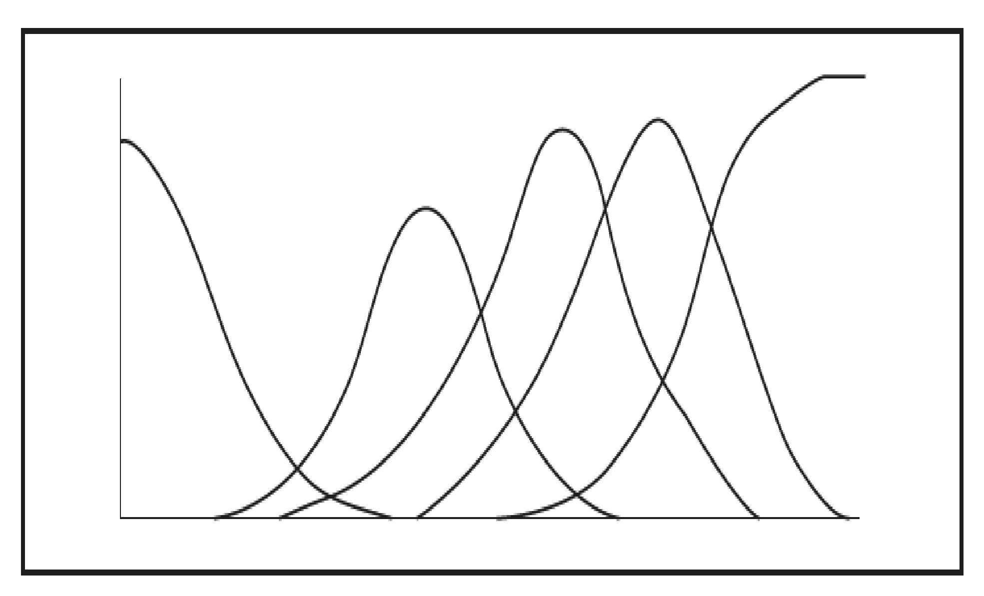
الطرق القائمة على نظرية الاختبار للبند IRT ستستمر شعبيتها في الزيادة. وسيكون لها مزايا واضحة على الطرق الســابقة في ظروف معينة. سوف تتعايش نظرية الاستجابة للبند والأساليب الكلاسيكية، بقدر مشاركة تحليل الانحدار للمرحلة مع أساليب نمذجة المعادلات البنائية. على الرغم من أن كلاً من نظرية الاستجابة للبند ونمذجة المعادلات البنائية حملتا أشياء أبعد من سوابقهما، إلا أن الطرق السابقة احتفظت بفائدتها.

تمارين:

١- تمثل الأشــكال التالية نتائج تحليل نظرية الاســتجابة للبند لاثنين من البنود الافتراضية، ولكل منها خمسة خيارات استجابة.



شكل (٧-١٢): المنحنى المميز للبند (ICC) الافتراضي١



شكل (٧-١٣): المنحنى المميز للبند (ICC) الافتراضي٢

١أ - بنــاءً عــلى تلك الأشــكال البيانية، قم بتقييــم البنود فيما يتعلق بما يــلي واشرح ما هي خصائص المنحنى كأساس لتقريرك.

١ب - أي من البندين يقوم بعمل أفضل في التقاط النطاق الكامل للظاهرة التي يقيّمها؟

١ج - أي من خيارات الاســتجابة على البند تؤدي وظيفة أفضل في التمييز بين المســتويات المختلفة للظاهرة؟

٢- صِــف الاختلاف في المظهر للمنحنيات المميزة للبنود (ICCs) لبندين ثنائيي الاســتجابة binary items مختلفين في معايير التمييز الخاصة بهما.

الفصل الثامن
القياس في السياق البحثي الأوسع

يمهد الفصل الافتتاحي لهذا الكتاب المجال لما يجب متابعته من خلال تقديم بعض الأمثلة عن ومتى ولماذا تبرز مشكلات القياس، ومناقشة دور النظرية في القياس، والتأكيد على الاقتصاد الخاطئ جراء العجلة في إجراءات القياس. وبشكل رئيسي، رسم الفصل الافتتاحي السياق البحثي الأوسع قبل أن ينتقل التركيز إلى القضايا المحددة التي تتناولها الفصول اللاحقة. هذا الفصل يعود إلى الصورة الكبيرة وينظر بإيجاز إلى المقياس ضمن السياق الأكبر للبحث.

قبل تطوير المقياس:

ابحث عن الأدوات الحالية:

في بدايات هذا الكتاب، اقترحت أن تطوير المقياس ينشأ غالبًا من نقص الأدوات الموجودة المناسبة. إن التأكد من عدم وجود بديل قياس مناسب مسبقاً يعتبر أمراً هاماً وفعالاً. وفي مكان آخر اقترح ديفيليس (DeVellis, 1996) طرقًا للبحث عن مقاييس مناسبة. بشكل أساسي، تتضمن هذه العملية البحث في ملخصات منشورة وإلكترونية للمقاييس لتحديد ما إذا كان هناك مقياس مناسب موجود مسبقاً. وتحتوي السلاسل المنشورة مثل الكتاب السنوي للقياسات الذهنية Mental Measurements Yearbook (مثلاً سبيس، كارلسون، وجينسينجر Spies, Carlson, & Geisinger, 2010) واختبارات في الطباعة Tests in Print (Murphy, Spies, & Plake, 2006) على مقاييس سريرية في المقام الأول، بما في ذلك اختبارات القدرات والشخصية. غالبًا ما تكون هذه هي الأدوات التي يستخدمها علماء النفس التطبيقيون لتقييم العملاء. الأدوات المخصصة بشكل أساسي للبحث تكون أقل شهرة، ولكن يتم أخذ بعضها في الاعتبار. هناك فئة أخرى من المصادر هي المجموعات المستهدفة، مثل المقاييس الكلاسيكية للشخصية Measures of Personality

ومقاييس المواقف النفسـية الاجتماعية Social Psychological Attitudes (روبينسـون، شيفر، فيرجيتسمان، Robinson, Shaver, & Wrightsman,1991). كما أن المجلات ذات الصلة هي أيضًا مكان ممتاز للعثور على إسـتراتيجيات القياس التي حققت نجاحاً للآخرين المهتمين بنفس المفهوم البنائي محل الدراسة.

مـع التكرار المتزايـد، يمكن للباحثين العثـور على معلومات حـول أدوات القياس على الإنترنت. في الواقع، قد يكون الإنترنت هو المكان الذي يظهر فيه التوسع السريع في المعلومات المتعلقة بالقياس. وكمثال على التوسـع في مصادر الإنترنت، أصبح كل من الكتاب السنوي للقياسات الذهنية (Murphy et al., 2006) والاختبارات في الطباعة (Spies et al.,2010) قابلاً للبحث عبر الإنترنت في موقع الويب الخاص بمعهد بيروز Buros للقياسات العقلية في جامعة نبراسـكا في لينكولن (http://www.unl.edu/buros/). مثال آخر بارز لمصدر من المصادر على الإنترنت للحصول على معلومات حول المقاييس، هو موقع الويب المتعلق بنظام معلومـات قياس نتائج تقارير المـرضى Patient-Reported Outcomes Measurement Information System (PROMIS). نظـام معلومـات القيـاس لنتائـج تقارير المرضى PROMIS هـو مبادرة بحثيـة تدعمها المعاهد الوطنية للصحة كجـزء من خارطة طريق البحـث الطبي (انظر ادر، Ader,2007، للحصول على نظرة عامة) لتحقيق ترتيب ووضوح أكبر لتقييم الحـالات النهائية المهمة في البحوث الصحية (PROMIS, 2010). هذا الموقع هـو بوابـة للمعلومات التي تم تطويرهـا من خلال جهد تعاوني يضـم جامعات متعددة وترعـاه المعاهد الوطنية للصحة. قام محققـو نظام PROMIS بمراجعة وتقييم وتصنيف وتعديل وتطوير بنود تقييم النتائج الصحية في خمسة مجالات واسعة: الأداء البدني، الأداء الاجتماعـي، الضيق العاطفي، الألم، والإجهاد. ولقد خضعـت هذه البنود للفحص الدقيق باسـتخدام طرق تسـتند إلى نظرية الاسـتجابة للبند IRT. البنود الموجودة في بنوك تلك المجالات المختلفة قد تمثل أكثر البنود التي تمت دراستها بعناية. ومع استمرار جهود نظام PROMIS، سـتخضع البنود للتحقق من الصحة على نطاق واسـع ومـن المحتمل إضافة مجموعات بنود جديدة.

على الرغم من وجود اسـتثناءات، مثـل تلك المذكورة أعلاه، تميل مسـتودعات أدوات القياس القائمة على الإنترنت والمستخدمة على نطاق واسع إلى أن تكون مؤقتة. إن متطلبات

التمويل والتوظيف والقضايا الإدارية الأخرى تعقد صيانة مثل هذه المواقع، ونتيجة لذلك، لا يدوم الكثير منها لأكثر من بضع سنوات. ومع ذلك، لا يزال الإنترنت مصدراً قيماً للباحثين الذين يسعون إلى إيجاد المقاييس. تعد محركات البحث على الإنترنت أداة ممتازة لتحديد أدوات القياس التي تستهدف متغيرات معينة. بمعنى أن التطورات في خوارزميات البحث أدت إلى تقليل الحاجة إلى المستودعات الموضوعة على المواقع. فإذا تم تطوير أداة لقياس متغير معين، فمن المرجح أن يجد البحث على الإنترنت إما الأداة نفسها أو معلومات حول كيفية تحديد موقعها.

كما هو الحال مع أي معلومات تستند إلى الإنترنت، يحتاج المستهلك إلى النظر في أصل المعلومات ومصداقيتها. فعادةً ما تقدم المواقع التي ترعاها الجامعات أو الهيئات الحكومية أو المؤسسات أو المنظمات الأخرى المنشأة معلومات دقيقة وذات مصداقية. بشكل عام، ومع ذلك، هناك ما يبرر الحذر. مثلما توجد وفرة من كتب «العلوم التافهة»، هناك أيضًا الكثير من مواقع الإنترنت التي قد تتبنَّى أسلوب ومظهر الشرعية العلمية التي لا تتناسب مع محتواها. يجب أن تساعدك المهارات التي اكتسبتها من هذا الكتاب في تقييم مصادر معلومات القياس بشكل ناقد وبأي صيغة وتحديد ما إذا كانت المقاييس الموضحة قد أظهرت ثباتاً وصدقاً كافياً. ويعتبر ما إذا كان المقياس قد تم نشره في مجلة علمية محكمة أم لا، وسيلة أخرى لتقييم موثوقيتها. بالطبع، بالإضافة إلى اجتياز المقياس للتحكيم العلمي، لن يكون مفيدًا إلا إذا كان مناسبًا لتصور الباحث للمتغير محل الاهتمام. وبالتالي، هناك ما يبرر تمحيص مقياسٍ معينٍ بعد النشر في مجلة ما قبل تقرير أنه يعد اختيارًا جيدًا لمشروع بحث معين.

عرض المفهوم البنائي في سياق المجتمع محل الاهتمام:

غالبًا ما يكون من المهم تقييم ما إذا كانت المفاهيم البنائية النظرية theoretical constructs التي نحددها نحن الباحثين تتوافق مع التصورات والخبرات الفعلية للأشخاص الذين نخطط لدراستهم. يمكن أن تكون مجموعات التركيز (مثل Krueger & Casey, 2000) والأساليب النوعية الأخرى (على سبيل المثال، دي والت، روسروك، يونت، ستون DeWalt, Rothrock, Yount, & Stone, 2007) وسيلة لتحديد ما

إذا كانت الأفكار التي تقوم عليها المفاهيم البنائية محل الاهتمام منطقية بالنسبة للمجيبين. كمثال، خذ بعين الاعتبار السمات - التفسيرات أو التأويلات التي ينسبها الأشخاص إلى نتائج مختلفة. غالبًا ما يتم تقييمها وفقًا لأبعاد مثل "تحت سيطرتي" مقابل "ليس تحت سيطرتي" أو "خاص بهذا الموقف" مقابل "ينطبق على معظم الحالات" أو "خاصيتي" مقابل "خاصية البيئة أو الموقف". كانت البحوث في عمليات معالجة السـمات مثمرة للغاية، حيث يمكن لمعظـم النـاس تحليل النتائج، مثل تلقي عرض عمل بعد المقابلـة، على ضوء هذه الأبعاد. في بعض الحالات، قد لا يكون هذا هو الحال. على سـبيل المثال، إن مطالبة كبار السن من غير المولودين في المناطق الريفية والمقيمين في المناطق الريفية والذين يفتقرون إلى التعليم، والذيـن لا يعتـادون على التفكير بمثل هـذه المصطلحات، بتقييم نتائـج المرض أو قرارات الشراء وفقًا لهذه الأبعاد الثلاثة قد لا تعمل بشـكل جيد حيث تشـير التجربة إلى أنهم قد لا يفهمون المهمة ببساطة لأنها غريبة على طريقة تفكيرهم في الأشياء. إن مجموعة التركيز التي تطلب من المشاركين المحتملين في البحث مناقشة المفاهيم ذات الصلة قد تجعل هذه المشكلة واضحة وتمنع إستراتيجية القياس المحكوم عليها بالفشل.

يمكن لمجموعات التركيز أيضًا أن تكشـف عن اللغة الطبيعية اليومية التي يسـتخدمها الناس للحديث عن مفهوم ما. فقد لا تستخدم الأم الشابة نفس المصطلحات التي يستخدمها خبير التسويق لوصف ردود الفعل على المنتج. وقد تستخدم الأم الشابة «التظاهر» لوصف طفلها يلعب دون اسـتخدام لعبة معينة، بينما قد يصف باحث تسـويق مثل هذا الترويج بأنه «منتج غير موجه». بناء البنود بما يتناسب مع لغة الأم المستخدمة (على سبيل المثال، «كـم مقدار الوقت من اللعب الـذي يقضيه طفلك متظاهراً باللعب، دون اسـتخدام أي ألعاب؟») بدلاً من اسـتخدام لغة الأم الخبيرة (على سـبيل المثـال، «كم من الوقت يقضيه طفلك في اللعب غير الموجه نحو المنتج؟») من المرجح أن ينتج أداة مناسبة لقياس تصوراتها لكيفية تفاعل طفلها مع مختلف المنتجات.

المقابـلات المعرفية هي وسـيلة نوعية أخرى لتحديد كيفية تفسـير المجيبين المحتملين للبنـود وفهمها. على الرغم من وجود مجموعة متنوعة من الأسـاليب (على سـبيل المثال ويليز، Willis, 2005)، فإن الفكرة الأساسـية هي التعلم من المجيبين عن طريق سـؤالهم عما فهموه حول عنصر ما وكيف صاغوا اسـتجابة له. يمكن أن تكشف هذه الطريقة غالبًا

عن التشـويش حول المفردات أو المفاهيم أو سـوء الفهم المتعلق بخيارات الاستجابة التي قـد يتجاهلها الباحث دون إجراء مقابلات معرفية. يوفر هذا الأسـلوب ضمانًا إضافيًا بأن الباحث والمسـتجيب لديهم فهم مشـترك لمعنى البنود، وبالتالي، يمكـن أن يعزز ذلك من صدق المقياس.

وكملحوظة تحذيرية: يدعو بعض الباحثين إلى جعل المجتمع المستهدف يعطي الموافقة النهائية على الاستبيانات. هذا أمر مثير للإعجاب ومن المرجح أن يعطي المشاركين إحساسًا أكـبر بالمشـاركة الفعالة في عملية البحـث؛ ومع ذلك، من المهم التعـرف على حدود خبرة المجيبين. قد يعرفون أفضل من الباحث كيف يتحدثون هم وأشخاص من خلفيات متشابهة حـول قضايـا مختلفة، وأنهم مؤهلون بشـكل فريد لتقديم رؤى حـول فهمهم لمحتويات الاسـتبيان. يمكن لمجموعات التركيز والمقابلات المعرفية الاسـتفادة من الخبرة التي يتمتع بها أفراد المجموعة المسـتهدفة فيما يتعلق بهذه الأمور. لكنهم ليسـوا خبراء في التفاصيل الفنيـة لبناء المقياس. ليـس من العدل أن نتوقع من غير الخـبراء أن يفهموا القضايا التي تمت مناقشـتها في الفصل الخامس، على سبيل المثال، لأنها تنطبق على بناء البنود. وكمثال على ذلك، قد يفضل شـخص غير خبير بنداً تمـت صياغته بطريقة جذابة ومعتدلة، في حين أن مطور المقياس ذا الخبرة قد يدرك أن هذه الصياغة سـتولد تباينًا بسيطًا في الاستجابات، مـما يجعل البند غير ذي فائدة. توصيتي الشـخصية هي: أن تكون حساسًـا لاهتمامات المشـاركين، وساعدهم على المشـاركة بفعالية في مجموعة متنوعة من الطرق إذا كان ذلك مناسبًا، والاسـتفادة الكاملة من خبراتهم فيما يتعلق بالجوانب اللغوية أو الثقافية لصياغة البنود، واسـتمع إلى ملاحظاتهم بشأن المعنى من البنود للمجموعات ذات الاهتمام، ولكن احتفظ بحق الموافقة النهائية على صياغة البند. نحن لا نكرم المشـاركين في بحثنا إذا قمنا عـن غير قصـد بخلق موقف لا يمكن فيـه قياس وجهات نظرهم أو مشـاعرهم أو آرائهم بدقة؛ نحن ببسـاطة نضيع وقتهم. الهدف هو فهم ما هي طرق التعبير عن المفاهيم التي سـتكون أكثر وضوحًا للمستجيبين مع الحفاظ على سلامة تلك المفاهيم. القيام بذلك يمكن أن يستفيد من فهم الرؤى الخاصة التي يقدمها المشاركون والباحثون والاستفادة من نقاط القوة في مجموعتي المشاركين والباحثين.

اتخاذ القرار بشأن طريقة تطبيق المقياس:

يمكــن للباحثــين جمــع البيانــات بعــدة طــرق (مثــل ديلــمان، Dillman, 2007؛ فاولر Fowler, 2009)، وقد يختارون طرق تطبيق المقياس بناءً على تفضيلات المجيبين. ووفقًا لذلــك، قد يفكر فريق من الباحثين في اســتخدام طريقة المقابلة بدلاً من طريقة الاســتبيان المطبــوع. ومــن المهم أن ندرك أن المقيــاس المراد إكماله في النمــوذج المطبوع قد يكون له خصائص مختلفة اختلافًا كبيرًا عند عرض البنود والردود شــفهياً. على سبيل المثال، قد يكون الآباء أكثر ترددًا في الاعتراف بالطموحات العالية إذا كان عليهم الإبلاغ عنها بصوت عالٍ لمقابلة بدلاً من الاستبيان المطبوع. بشكلٍ عام، من الحكمة قصر أسلوب تطبيق مقياس جديد على الطريقة المســتخدمة أثنــاء تطوير المقياس. وكبديل لذلك، يجب عــلى الباحثين فحص تأثير أسلوب تطبيق المقياس بشكل منتظم. ويمكن استخدام دراسة التعميم (انظر الفصل الثالث) لتحديد مدى إمكانية تعميم المقياس من خلال طرق تطبيق المقياس المختلفة.

الأخذ في الاعتبار المقياس في سياق مقاييس أو إجراءات أخرى:

ما هي الأســئلة أو إجراءات البحث التي ستســبق المقياس نفســه؟ كيف ســتؤثر هذه الأســئلة على الإجابات عــلى المقياس؟ يشــير نونــالي (Nunnally, 1978, pp. 627-677)، إلى عوامل ســياقية مثل أنماط الاســتجابة، والتعب، والدافع كمتغيرات مؤثرة contingent variables. ويشــير أيضًا إلى أنه يمكنها التأثير سلبًا على الأبحاث من خلال ثلاث طرق: (١) عــن طريق تقليل ثبــات المقاييس؛ (٢) من خــلال تكوين مصادر موثوقــة للتباين بخلاف المفهــوم البنائي محل الاهتمام، وبالتالي تقليل الصــدق؛ و(٣) عن طريق تغيير العلاقات بين المتغيرات، وجعلها تبدو، على ســبيل المثال، أكــثر ارتباطًا مما هي عليه بالفعل. كمثال على كيفيــة عمــل المتغيرات المؤثرة، فكر في تحفيز المزاج والمجموعــات المعرفية لأنها قد تنطبق على مثال أبحاث التســويق. قد يشــكل تحفيز المزاج مشكلة إذا قرر الباحثون في التسويق، على ســبيل المثال، تضمــين مقياس الاكتئــاب أو تقدير الذات في نفس الاســتبيان كمقياس لتطلعاتهــم. غالبًا ما تحتــوي المقاييس التي تتطرق لهذه المفاهيــم البنائية (وغيرها) على بنــود تعبر عن آراء ســلبية عن الــذات. إن مقيــاس روزنــبرغ Rosenberg لتقدير الذات (Rosenberg, 1965)، عــلى ســبيل المثــال، يحتــوي على مثل هــذه البنود الســلبية مثل "أشــعر أنــه ليس لدي الكثير لأفتخر بــه" (بالإضافة إلى بنود تعبر عــن تقييم ذاتي إيجابي).

يمكن للباحث الذي لم يكن حساسًـا للآثار المحتملة لاسـتقراء الحالة المزاجية أن يختار حصرياً سلسـلة من البنود ذاتية النقد وضمها إلى مقياس تم تطويره حديثًا. قد تؤدي قراءة العبارات التي تعبر باستمرار عن التقييمات السلبية لنفس الشخص إلى حدوث حالة خلل يمكن أن تتسبب بدورها في إدراك ما هو آت بشكل مختلف عما قد يحدث (على سبيل المثال كيهلستروم، إريك، ساندبراند، ووتوبايس Kihlstrom, Eich, Sandbrand, & Tobias, 2000؛ رهليس، ريسكند، لان، Rholes, Riskind,& Lane, 1987). هذا قد يؤدي إلى كل من الآثار السلبية الثلاثة التي لاحظهـا «نونالي» Nunnally (1978). وهذا يعني أنه في حالة وجود بنود سـلبية مؤثرة، قد تأخذ بنود الطموح درجة مختلفة من المعنى، مما يقلل نسبة التباين في تلك البنود التي يمكن عزوهـا إلى المتغـير الكامن المقصود. أو، في الحالة القصوى، قـد تتأثر بعض البنود من مقياس الطموح بشـكل أسـاسي بحالة الحالة المزاجية المسـتحثة، مما يجعل المقياس متعدد العوامل ويخفـض صدقه كمقياس لطموح الوالدين. أخيرًا، إلى المدى الذي يؤثر فيه مزاج المجيبين على اسـتجاباتهم لبنود الطموح، قد ترتبط درجات هذا المقياس ارتباطًا مصطنعًا بالمقاييس الأخرى ذات الصلة بالمزاج.

المجموعـات المعرفية هي مثال أكثر عمومية لنفس الظاهرة. بمعنى أنه قد يتم تحفيز إطـار مرجعي مـا بخلاف المزاج من خلال تركيز انتباه المجيبـين على موضوع محدد. على سـبيل المثـال، قبل تطبيق مقياس الطمـوح مباشرة، فإن بنوداً تتعلق بدخل المسـتجيبين، وقيمة منازلهم، والمبالغ التي ينفقونها سـنويًا على فئات مختلفة من السـلع الاسـتهلاكية، قـد تخلق حالة ذهنية تعمل على تغيير طموحاتهم تجاه أطفالهم مؤقتًا. نتيجة لذلك، قد تعكس الاسـتجابات للمقياس حالة عابرة غير مقصودة. كما هو الحال مع الحالة المزاجية، قـد تؤثر هذه المجموعة المعرفية سـلبًا على ثبات و/ أو صـدق المقياس عن طريق تلويث المدى الذي يعكس فيه بشكل لا لبس فيه طموحات الوالدين.

بعد تطبيق المقياس:

تظهر مجموعة مختلفة تمامًا من القضايا بعد اسـتخدام وتطبيق المقياس لمعالجة سؤال بحثي موضوعي. وتبقى الاهتمامات الرئيسية في هذه المرحلة هي تحليل وتفسير البيانات التي تم توليدها من خلال الأداة.

قضايا تحليلية:

إحدى القضايا في تحليل البيانات هي مدى ملاءمة التقنيات المختلفة للمتغيرات ذات خصائص القياس المختلفة. لذا؛ فإن المنظور النظري والأساليب التي نوصي بها بقوة في هذا الكتاب يجب أن ينتج عنها مقاييس قابلة لمجموعة واسعة من طرق تحليل البيانات. على الرغم من أن، بالمعنى الحرفي للكلمة، البنود التي تستخدم صيغ ليكرت Likert أو الاستجابة التفاضلية الدلالية semantic differential response قد تكون رتبية، إلا أن الخبرة الثرية المتراكمة تدعم تطبيق الأساليب التحليلية المستندة إلى القياس الفتري interval-based analytic methods على المقاييس التي تنتجها. ومع ذلك، فإن السؤال كان حول ما هي الأساليب الأنسب لنوع البيانات، وبالتأكيد سوف يستمر، في نقاش ساخن في العلوم الاجتماعية. ويعد تحديد كيفية تأثير خيارات الاستجابة المختلفة على تقديرات المتغيرات الأساسية مجالًا نشطًا للبحث في حد ذاته. أيضاً، سيكون لدى مختلف الجماهير توقعات مختلفة لكيفية التعامل مع المقاييس. وفي حين أن علماء النفس، على سبيل المثال، قد يكونون متفائلين إلى حد ما في التعامل مع قياسات ليكرت على أنها تنتج بيانات على المستوى الفتري، فإن علماء الأوبئة قد لا يكونون كذلك. ربما يكون النهج الأكثر عملية هو المراقبة (والتوافق) مع التوجه السائد فيما يتعلق بهذه القضية في مجال اهتمام الفرد.

قضايا التفسير:

على افتراض أن الباحث قد توصل إلى خطة تحليلية مناسبة للبيانات التي تم إنشاؤها بواسطة مقياس تم تطويره حديثًا، فإن مسألة كيفية تفسير البيانات لا تزال قائمة. إحدى النقاط التي يجب وضعها في الاعتبار عند هذا المنعطف هي أن صدق المقياس لم يتم تأسيسه بشكل قاطع أثناء تطوير المقياس. التحقق من الصدق عملية تراكمية مستمرة. علاوة على ذلك، فإن صدق المقياس هو في الحقيقة إحدى خصائص كيفية استخدام المقياس، وليس المقياس نفسه. مقياس الاكتئاب، على سبيل المثال، قد يكون صادقاً لتقييم الاكتئاب ولكن ليس لتقييم التأثير السلبي العام.

أيضًا، من المهم التفكير في النتائج التي تم الحصول عليها. خاصةً إذا كانت النتائج تبدو غـير بديهية أو مضادة للنظرية، فيجب على الباحث عندئـذ أن يفكر في احتمال أن يكون المقياس غير صادق في سياق تلك الدراسة بالذات (إن لم يكن على نطاق أوسع). قد يكون مـدى صدق المقياس الذي يمكن أن يعمم على المجتمعـات أو الأطر أو التفاصيل المحددة لتطبيق المقياس أو مجموعة متنوعة من الأبعاد الأخرى محدوداً. على سبيل المثال، مقياس طمـوح الوالديـن الافتراضي ربما يكون قد تـم تطويره مع وجود مجموعة غنية نسـبياً في الاعتبار، وقد يكون من غير المقبول صلاحيته للأفراد الذين تكون مواردهم محدودة. يجب أن تأخذ أي استنتاجات تستند إلى مقياس محدود الاستخدام بعين الاعتبار ما يلي: (أ) كيف يختلف تطبيقه الحالي عن سـياق الصلاحية الأصلي الخاص به، (ب) احتمال أن تؤدي هذه الاختلافات إلى الحد من صدق المقياس، و(ج) آثار تلك القيود على البحث الحالي.

قابلية التعميم:

تم التحذير في الفقرة السـابقة من التعميم عـلى المجتمعات والأطر والجوانب الأخرى للبحث. هذه القضية تستدعي المزيد من التركيز؛ فالتوصل إلى استنتاجات حول الفروق بين المجموعات قد يتداخل مع الاختلافات في الظاهرة التي يتم قياسها والاختلافات في أداء الأداة. إذا اسـتطعنا أن نفترض أن الأخير (أي، الاختلافات في الظاهرة التي يتم قياسها والاختلافات في أداء الأداة) طفيف، فيمكننا أن نسند الفروق الملحوظة إلى عضوية المجموعة. في العديد من المواقف (على سـبيل المثال، مقارنة وقت الانتهاء من المهمة في مجموعتين من الأطفال الذين تم اختيارهم وتخصيصهم بشـكل عشـوائي للمجموعتين)، ستكون هذه هي الحالة. في مجموعات أخرى (على سبيل المثال، المقارنات بين المجموعات المتمايزة ثقافيًا)، لا يمكننا افـتراض أداء متطابق للأداة. يعد أداء البنـد التفاضلي **differential item functioning**، الذي تمت مناقشـته باختصار في الفصل السابع، مجالًا نشـطًا للبحث السيكومتري. وعلى الرغـم من أن معظم الباحثين لن يجعلوا من تلك القضايا محورًا لجهودهم الخاصة، إلا أنه ينبغي عليهم إدراك إمكانية وجودها والقيود التي قد تفرضها على استنتاجاتهم.

أفكار ختامية:

القياس الصغير والقياس الكبير:

الجانـب الأخير مـن القياس الذي يسـتدعي النظر هو نطاق المهمـة وتعقيد الأدوات اللازمة لها. في مكان آخر من هذا الكتاب (على سـبيل المثال، في مناقشـة صدق المحتوى)، تحدثـت عن مدى اتسـاع المفهوم البنائي construct الذي صُممـت الأداة لتقييمه. وأود العـودة هنـا إلى موضوع ذي صلة مـن خلال النظر في كيفية ارتباط نطاق سـياق القياس بتعقيد مجمل جهد القياس. هنا سأبدأ بتشبيه أو محاكاة.

الزوارق والسفن البحرية:

عندما كنت طالب دراسـات عليا، أمضيت فترة ما بعد الظهر في مسـاعدة صديق كان يبني زورقاً من مجموعة أدوات اشتراها عبر البريد. لقد كانت هذه مهمة مجدية اقتصاديًا وتكنولوجيًـا لشـخص ملتزم بـكل النفقات المعتـادة والمتطلبات الزمنيـة لحضور برنامج الدراسـات العليا، وقد خدم هذا الزورق صديقي وعائلته جيدًا لسـنوات عديدة. لو كان قد اقترح عليّ أن أمد له يد العون باستخدام بعض الأموال التي يمكن ادخارها لبناء سفينة بحريـة، لربما أحلته إلى مرفق الصحة العقلية بالحـرم الجامعي. لكنه لم يكن في حاجة إلى سـفينة بحريـة لأنواع رحلات المياه التي تَصوّرَها وتمتع بهـا في نهاية المطاف. كان الزورق عمليًا ومناسبًا للاستخدامات التي قصدها.

الزوارق والسـفن البحرية على حد سواء مراكب قيّمة على الرغم من اختلافها الواضح. ربمـا يكون الاختلافان الأكثر وضوحًا بينهما حجمهما ووظيفتهما. الزوارق عبارة عن قوارب صغـيرة ذات قـدرة عالية على المناورة مثالية للاسـتخدام في المياه المغلقة نسـبيًا من قِبل شخصين أو ثلاثة. إذا كنت ترغب في استكشاف نهر لطيف، واستغراق بعض الوقت لدخول مختلف الخلجان الصغيرة وفحص المواطن الطبيعية بشيء من التفصيل، فإن الزورق المصمم جيدًا والمصمم بعناية يعد وسيلة مناسبة للغاية.

سفن الرحلات البحرية هي سفن كبيرة جدًا قادرة على استيعاب آلاف الركاب وتزويدهم بمجموعة من وسائل الراحة والأنشطة أثناء عبور البحار المفتوحة. لقد تمت مقارنتها بالمدن

العائمة فيما يتعلق بمجموعة الخدمات والمرافق التي تقدمها. وتشـمل هذه عادةً قاعات للاحتفالات، ومرافق طعام شاملة، وأماكن إقامة مريحة للنوم، وخدمات متخصصة يقدمها فريق كبير من الموظفين. تعتبر سـفينة الرحلات البحرية أكثر من وسـيلة مواصلات، فهي تؤدي وظائف متعددة، مثل الترفيه والإقامة، ويمكنها تقديم هذه الوظائف بشكل جيد على قدم المسـاواة في أجزاء مختلفة من العالم. الحجم والوظيفة ليستا الاختلافات الوحيدة بين هذين النوعين من القوارب المائية. تشغيل كل نوع يتطلب مستويات مختلفة من الخبرة. فمعظم البالغـين الأصحاء (والكثير من الأطفال) الذين لديهم بعض الإرشـادات والخبرات القصيرة نسـبيًا يمكنهم تجربة الزورق. من ناحية أخرى، تتطلب سـفينة الرحلات البحرية طاقـمًا مدربًا ومعتمدًا من المهنيين البحريين لتشـغيلها بأمان. يمثل الوقت اللازم لتصميم الزورق وبناءه وتشغيله مجرد جزء بسيط من النطاق الزمني مقارنةً بجلب سفينة سياحية للخدمة. النماذج التي تتخذها الزوارق وسـفن الرحلات البحرية يتم استنباطها من خلال قواعـد المعرفة المختلفة، مع تـوازن مختلف تماماً بين الخـبرة والتكنولوجيا المتقدمة التي تميزها. ربما بسـبب بساطتها النسبية، كانت الزوارق بأشكال مختلفة مستخدمة منذ آلاف السـنين. وبالتالي، لدينا ثروة من المعرفة العملية عنهم والتي يمكن أن تسـاعدنا على فهم نقاط القوة والضعف في بدائل التصميم المختلفة. بالمقابل، تتمتع سـفن الرحلات البحرية بتاريخ أقصر بكثير ولكنها تستفيد من التطورات التكنولوجية في التصميم والتصنيع. أخيرًا، في حين أن تكلفة الزورق متواضعة نسـبيًا، فإن تكلفة تصميم وبناء وتجهيز وتزويد سفينة الرحلات البحرية أكبر بعدة مرات.

تقوم كل من الزوارق وسـفن الرحلات البحرية بعملها بشـكل جيد. تعد الزوارق هي الأفضل في حالة الرحلات القصيرة، رغم أنه فيما مضى من الزمن، تم إكمال رحلات لمسافات طويلة بنجاح باستخدام الزوارق. ما نسميه الآن جزر هاواي، على سبيل المثال، من المحتمل أنه تم اسـتعمارها بدايةً من قبل المسـافرين مـن أجزاء أخرى مـن بولينيزيا حيث كانوا يسافرون مسافات طويلة عبر الزوارق. ومع ذلك، يفضل معظمنا اليوم عبور المياه المفتوحة لجنوب المحيط الهادئ على متن سـفينة سياحية، مع الرفاه والراحة والسلامة التي توفرها. على الرغم من أن الزورق كان بإمكانه (ويفترض أن يفعل) إنجاز المهمة، إلا أن سفينة أكثر تطـوراً وقدرة تبدو أكثر ملاءمة للوظيفة، والرحلات عبر بولينيزيا في هذه الأيام من المرجح

أن تكون باســتخدام ســفن الرحلات البحريــة أكثر من الزوارق. مــن ناحية أخرى، بمجرد وصول ســفينة الرحلات البحريــة إلى هاواي من أرخبيل بولينيزي آخر، قد يتنقل المســافر من ســفينة الرحلات ويســتخدم مركبة أصغر بكثير، مثل الزورق أو قوارب الكاياك، للقيام ببعض الاستكشــافات عن قرب في الجزر والممرات المائية الأصغر. لا يمكن ببساطة لسفينة الرحــلات الكبيرة المرهقة أن تعمل كزورق عندما يتم تضييق مقاس الممر المائي. أضف إلى ذلك أنه يمكن بســهولة نشر قوارب متعددة لاستكشــاف مناطق مختلفة على طول ساحل متعرج أو مجرى مائي دائري.

لذلك أيهما أفضل، ســفينة رحلات سياحية أم زورق؟ أود أن أقترح أن لكل منهما مكانه وغرضــه، وعلى الرغم من أن وظائفهما تتداخــل إلى حد ما، فمن الواضح أن هناك مواقف يفضل أحدهما على الآخر. من المؤكد أن الحصول على الزورق أقل تكلفة، ولكن على الرغم من الرحلات البطولية للزوارق التي اســتكملت في العصور القديمة، فمن المحتمل أن تكون مقصورة على مهام بســيطة. من ناحية أخرى، فإن التكلفة، والخبرة التكنولوجية، والوقت، والجهد العام الذي ينطوي عليه بناء سفينة سياحية، يحول دون استخدامها في نزهات مائية أكثر تقييدًا تكون فيها مركبة أصغر حجمًا كافية تمامًا.

القياس «الزوارق» والقياس «سفن الرحلات البحرية»:

بشــكلٍ عام، يمكن للمرء أن يقارن بين «القياس الصغير» و«القياس الكبير» ورسم بعض المقارنات لمناقشــتنا للقوارب المائية. ما أشــير إليه على أنه قياس صغير ينطوي على تصور وتطويــر وتقييــم المقاييس التي عادةً ما تكــون أكثر تركيزًا وتحديــداً. الأمثلة النموذجية هي الحالات العديدة التي يســعى فيها الباحث للإجابة، ضمن مســعى بحثي مقيد، على أســئلة محددة حول بعــض المتغيرات التي لا تتوفر لها مقاييس مناســبة. في هذه الحالة، يكرس الكثير من الباحثين الجهود لإنشــاء المقاييــس التي تتطلبها أبحاثهم. على الرغم من أن الظــروف ســتختلف بوضوح من مثال إلى مثــال للقياس الصغير، فــإن الباحث في كثير من الأحيان يتحكم بشــكل كبير في كيفية تشــغيل البناء على نطاق واســع أو ضيق. يمكن تصميم العناصر التي تم تطويرها لالتقاط البنية بشــكل متكرر لاســتهداف المتغير الكامن للاهتمام بدقة كبيرة ولتلبية بعض الافتراضات التي يمكن أن تبسط عملية التطوير والتحقق.

يمتد التطويــر الأولي في كثــير من الأحيان لعدة أشــهر وليس لعدة ســنوات، كما أن الموارد اللازمة والمهام التي ينطوي عليها هذا الجهد يمكن التحكم فيها نســبيًا. القياس الصغير، في هذا الصدد، يشبه الزورق في التشبيه السابق. وهذا يعني أن حجم المشروع بسيط نسبيًا، وســتكون وظيفة الأداة المطورة حديثًا محددة تمامًا، ومستوى الخبرة اللازمة لتصميم الأداة في متناول أيدي العديد من الباحثين الأكفاء، والوقت المطلوب في متناول اليد، والإجراء يمكن أن يســتفيد من الخبرة العملية بدلاً من الاعتماد على أحدث التطورات، والتكلفة الإجمالية ليست باهظة.

أعني بالقياس الكبير، العمل الذي يهدف إلى تقييم مفاهيم واسعة، ومتعددة الجوانب بطــرق تهــدف إلى تقديم نتائــج ومجموعــات أدوات قابلة للتعميم إلى حــد كبير ويمكن اســتخدامها على نطاق واســع عبر مجموعة من الظروف المتنوعــة والتي يحتمل أن تكون صعبة. بالإضافة إلى كون المفاهيم نفســها واســعة، فإن مســتوى الخبرة المطلوبة لتحقيق قياس كبير حقاً هو في الغالب أمر جوهري. وغالبًا ما تظهر للعيان مثل هذه الأنشطة خلال إطار زمني ممتد، وقد يعتمد تصميم الأدوات بشكل كبير على التطورات المتقدمة في نظرية القياس والممارســة أكثر من الخبرة العملية مع الأســاليب المجربة والحقيقية. نتيجة لهذه المطالب المختلفة، فإن التكلفة الإجمالية للقياس الكبير عادةً ما تكون كبيرة.

تشــمل الأمثلة على القيــاس الكبير الاختبارات التعليمية الموحــدة مثل اختبار القدرات الدراسية (SAT) واختبارات الخريجين القياسية (GRE) أو مشاريع التقييم الصحي الشاملة مثــل مبــادرة PROMIS. في هذه الحــالات، المتغيرات التي يتم قياســها معقدة وتتضمن مجموعة من المفاهيم الأكثر تحديدًا. تشــمل القدرات الدراســية، على سبيل المثال، سمات أكــثر تمييزًا مثل فهم القراءة والمفــردات والكفاءات الرياضيــة. وبالمثل مبادرة PROMIS واســعة النطاق، تضع مقاييس لمجموعــة متنوعة من النتائج المتعلقــة بالصحة تتراوح من الاكتئــاب إلى المشــاركة الاجتماعية إلى قيــود التنقل. هذه الأمثلة هــي أيضًا هرمية حيث تحــدد متغيرات محددة أكثر (مثل المهــارات النحوية أو قيود التنقــل) كجزء من المفاهيم الأوسع (على سبيل المثال، الاستعداد اللفظي أو الأداء البدني). التقييمات الشاملة للمفاهيم الواســعة مثل هذه النتائــج الأكاديمية والصحية تعتبر أنشــطة كثيفة الاســتخدام للموارد، وعادةً ما تشــمل ســنوات عديدة من الجهد وفــرق البحث الكبيرة ذات المهــارات العالية.

وكمثــال على الخبرة المتخصصة اللازمــة للجهود المبذولة في هذا المجال، عندما بدأت مبادرة PROMIS، كانــت تتكون من مراكز بحث متعددة تتبــع جوانب مختلفة من نطاق النتائج الصحيــة الأكــبر. يضمن كل فريق من هــذه الفرق واحدًا أو أكثر مــن المهنيين ذوي الخبرة الكبيرة في القياس. ومع تقدم العمل، واجه باحثو مبادرة PROMIS صعوباتٍ شكلت تحدياً للخبرات المتاحة للفرق الأساسية، وسعى البرنامج إلى الحصول على مدخلات إضافية من خبراء أكثر خبرة وتخصصًا. مع استمرار العمل إلى أبعد من ذلك، استعان الكادر الأصلي من الخبراء الخارجيين بآراء علماء القياس النفســيين الأكثر خبرة وتخصصًا. مشــاريع القياس الكبيرة مثل SATs أو PROMIS مماثلةً لسفن الرحلات البحرية متعددة الوظائف التي وصفتها سابقًا.

عــلى الرغم من أن أهداف ومتطلبات القياس الصغير عادةً ما تكون أقل من تلك اللازمة للقياس الكبير، فإن الكلمات الوصفية "الكبيرة" و"الصغيرة" لا تهدف إلى الإشارة إلى الاختلافات ذات الأهمية أو القيمة العلمية. في الواقع، العديد من مشــاريع القياس الكبيرة مســتوحاة من جهود القياس الصغيرة السابقة (تمامًا كما قد تتبعت سفن الرحلات البحرية اليوم الطرق التي كانت تعبرها الزوارق التي تعود إلى عصور ما قبل التاريخ). يمكن لمقاييس محددة ذات سمات مهمة أن تفرز برامج بحثية عالية الإنتاجية. وفي بعض الحالات، سيتم توسيع نطاقها، وقــد تنتقــل نماذج القياس المســتخدمة من قياس صغير إلى كبير. في حالات أخرى، ســتبقى المتغيرات المحددة للغاية محور اهتمام مهم دون أن تصبح أوســع أو تســتدعي الانتقال إلى محاولات قياس أكثر تفصيلاً أو واسعة.

الآثار العملية للقياس الصغير مقابل القياس الكبير:

غالبًــا ما تكون النماذج والأدوات التحليلية اللازمة للقياس الصغير أكثر بســاطة. بالطبع، يعتمد النموذج «الصحيح» والأســلوب التحليلي «الصحيح» بشكل أساسي على طبيعة أسئلة البحــث، ولكن مع القياس الصغير، قد يكــون الأمر أكثر ملاءمة لتصميم بنود تلبي افتراضات النماذج الأكثر بســاطة والتي تتطلب الأســاليب التحليلية الأقل تعقيداً. على سبيل المثال، في الفصل الخامس، عند مناقشة بنود مقياس ليكرت Likert scale، اقترحتُ معايرة صياغة البند من خلال تخيل كيف يمكن للمســتجيب العادي أن يجيب على كل مجموعة من البنود قيد الدراســة. وأوصيتُ أيضًا بتصميم البنود بحيث يحتمل أن يختار معظم المشــاركين استجابة

بالقرب من منتصف نطاق خيارات الاســتجابة المتاحة لهم. هذا يؤدي أساسًــا إلى الســعي للحصول على بنود ذات مؤشرات حساســة بنفس القدر لقوة مستوى كل مستجيب للمتغير الأساسي. والنتيجة الشائعة لهذه الخطوة البسيطة إلى حد ما هي أنها تميل إلى إنتاج بنود ذات تباينات وتشــبعات loadings على العامل المشــترك كبيرة نسبيًا ومتجانسة إلى حد ما. وهذا يعني أن البنود تميل إلى التوافق بشــكل أكبر مع افتراضات مكافئ «تاو» tau equivalence الأساسية. إذا كان الدليل التجريبي يكشف أن التباينات والتشبعات، في الواقع، متجانسة إلى حدٍ ما وأن البنود قد تم إنشــاؤها بشــكل صريح مع الأخذ في الاعتبار جعلها وبشكل متساوٍ مؤشرات جيدة للمتغير الأساسي، فإن الأساليب القائمة على مبدأ مكافئ «تاو» equivalency tau الأســاسي (مثــل معامل ألفا) قد تكون مناســبة. (دليلٌ آخر عــلى أن نموذج مكافئ تاو الأســاسيtau-equivalent قد يناســب البيانات، هو إنشــاء معامل ألفا المحسوب باستخدام إما التباين المشترك أو الصيغ القائمة على الارتباط، تعطي قيمًا متشابهة جدًا). أيضًا، محاولة كتابة بنود بحيث أن المستجيب العادي من المرجح أن يختار خيار استجابة بالقرب من مركز النطاق المتاح لخيارات الاستجابة سوف يميل إلى تقليل الالتواء skewness في البيانات. وكما كشــفت مناقشــة قيود ألفا alpha في الفصل الثالث، فإن تجنب الالتواء في البيانات ســوف يخفف من المشاكل المحتملة جراء استخدام ألفا عندما تكون خيارات الاستجابة فعليًا ليست من القياس الفتري interval scaled. وبالمثل، يمكن التبرير بســهولة بإعطاء أوزان متساوية للبنود عندما تكون التباينات والتشــبعات متجانسة نســبيًا بحيث يمكن لمجموع أو متوسط درجات البنود أن يخدم ببساطة كملخص للمقياس ككل.

إن التحكــم الأكبر الذي يتمتع به الباحث عــادةً في حالات القياس الصغير هو أن القياس الأصغر يتيح للباحث أســاليب أبســط من الناحية المفاهيمية والتحليليــة. قد يكون مفيدًا أيضًا، باســتخدام نفس البيانات، لمقارنة النتائج بين الأســاليب التقليدية (على ســبيل المثال، حســاب معامل ألفا استنادًا إلى ارتباطات بيرسون القياســية) والأساليب المطورة التي تستند إلى افتراضات أكثر ليبرالية (على سبيل المثال، حساب أوميغا كتقدير للثبات reliability). إن المقارنات من هذا النوع، باستخدام البيانات الحقيقية الناتجة من البنود التي تم إنشاؤها مع إدراك افتراضات النموذج الذي صممت البنود لتحقيقه، لا تزال نادرة نســبيًا وستقدم خدمة كبيرة لمجتمع البحث.

الأفــكار التي نوقشــت هنا تتعلق في المقام الأول بثبــات أداة القياس، وبطبيعة الحال، فــإن مدى ثبات المقياس يفرض قيودًا على صدقــه validity. ومع ذلك، كما هو مذكور في الفصل الرابع، يمكن أن يؤثر "التوســع الكبير" في بنود المقياس على الصدق بشــكل مباشر (على سبيل المثال، من خلال جعل كل أو بعض محتوى بنود المقياس غير صالح في سياقات معينــة). وبالتــالي، فإن نطاق جهد القيــاس المخطط مهم في مراحــل متعددة من تطوير المقياس ويستحق النظر فيه بعناية.

إذا كانت البيانات أو أسئلة البحث لا تبرر الأساليب الأبسط لتطوير المقياس، فيجب على الباحث اســتخدام كل ما هو ضروري لتقديم الأســباب الأوضح والأكثر دقة لعملية القياس والأداة الناتجة. في بعض الحالات، قد يتطلب ذلك البحث عن مشــاركة أو مســاهمة من متخصــصي القيــاس الأكثر خبرة. لمجرد أن المحيطات كانت قــد عُبِرت بالزوارق لا يعني أن القيام بذلك في وقتنا الحالي، عندما توجد سفن أكثر قوة، أمر جيد. ولكن على نفس المنوال، إذا أمكن تبسيط العملية بما يحقق الأهداف البحثية المقصودة بالقدر الكافي، فإنه يمكن أن تكون البســاطة الناتجة ذات فائدة كبيرة. بالإضافة إلى ذلك، الجهود والموارد غير الضرورية التي لا تفيد مرحلة القياس قد يتم تحويلها لصالح بعض الجوانب الأخرى للمشروع. وعلى الرغــم مــن وجود قيمة في فهم كيفيــة توظيف نماذج القياس والأســاليب التحليلية الأكثر تعقيــدًا، إلا أن هناك قيمة مكافئــة في إدراك - ويفضل أن يكون ذلك في مرحلة مبكرة من العمــل - أن توظيفهــا غير ضروري حقًا لتحقيــق الهدف المطروح. غالبًــا ما يوفر القياس الصغير فرصًا لمثل هذا التبسيط الذي نادرًا ما يحدث في القياس الكبير.

تذكَّر، القياس مسألة هامة:

القياس يمثل جانباً حيوياً في البحوث الاجتماعية والسلوكية. بغض النظر عن مدى جودة تصميــم وتنفيذ جوانب أخرى من المســعى البحثي، يمكن للقياس بناء أو هدم الدراســة. نفترض أن المتغيرات التي تهمنا تتوافق مع إجراءات التقييم التي نســتخدمها. في كثير من الأحيــان، العلاقة محل الاهتمام الرئيسي تكون بين اثنين أو أكثر من المتغيرات التي لا يمكن ملاحظتها أو رصدها، مثل الرغبة في الحصول على نتيجة معينة والفشل في التفكير في نتائج بديلــة. لا يمكننــا قياس الرغبة أو الاعتبار مباشرة، لذلك نبنــي مقاييس نأمل أن تتضمنها.

هذه المقاييس، هي إلى حدٍ ما، اســتعارات كمية للمفاهيم الأساســية. فقط إلى الحد الذي تكون فيه هذه الاســتعارات ملائمة (أي الأدوات صالحة)، فــإن العلاقات التي نلاحظها بين المقاييــس تعكس العلاقة التي نرغب في تقييمها بين المفاهيم البنائية غير القابلة للملاحظة أو الرصد. أخذ العينات الممثلة، وتصميم البحث المميز، والتنفيذ المتقن للإجراءات لن يغير هــذه الحقيقة. الباحث الذي لا يفهم العلاقة بين المقاييس والمتغيرات التي تمثلها، بالمعنى الحــرفي، لا يعــرف ما الذي يتحدث عنــه. إذا نظرنا إليها في ضوء ذلك، فــإن الجهود التي ينطوي عليها القياس الدقيق تتم مكافأتها بشكل وافٍ من خلال فوائدها.

تمارين:

١- صــف بإيجاز القياس الصغير والقياس الكبير، واذكر مزايا وعيوب كل منهما، وقدم أمثلة بحثية لكل منهما.

المراجع

- Ader, D. N. (2007). Developing the Patient-Reported Outcomes Measurement Information System (PROMIS). Medical Care, 45 (5, Suppl. 1), S1–S2.
- Ajzen, I., & Fishbein, M. (1980). Understanding attitudes and predicting behavior. Englewood Cliffs, NJ: Prentice Hall.
- Alder, K. (2002). The measure of all things: The seven-year odyssey and hidden error that transformed the world. New York: Free Press.
- Allen, M. J., & Yen, W. M. (1979). Introduction to measurement theory. Monterey, CA: Brooks/Cole.
- American Psychiatric Association. (1952). Diagnostic and statistical manual: Mental disorders. Washington, DC: Author.
- American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th ed., text revision). Washington, DC: Author.
- American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC: Author.
- Anastasi, A. (1968). Psychological testing (3rd ed.). New York: Macmillan.
- Aragona, M. (2009). The concept of mental disorder and the DSM-V. Dialogues in Philosophy, Mental, and Neuro Sciences, 2, 1–14.
- Asher, H. B. (1983). Causal modeling (2nd ed.). Sage university paper series on quantitative applications in the social sciences (Series No. 07-003). Beverly Hills, CA: Sage.
- Barnette, W. L. (1976). Readings in psychological tests and measurements (3rd ed.). Baltimore: Williams & Wilkins.

- Bjorner, J. B., Chang, C., Thissen, D., & Reeve, B. B. (2007). Developing tailored instruments: Item banking and computerized adaptive assessment. Quality of Life Research, 16, 95–108.

- Blalock, S. J., DeVellis, R. F., Brown, G. K., & Wallston, K. A. (1989). Validity of the Center for Epidemiological Studies Depression Scale in arthritis populations. Arthritis and Rheumatism, 32, 991–997.

- Bohrnstedt, O. W. (1969). A quick method for determining the reliability and validity of multiple-item scales. American Sociological Review, 34, 542–548.

- Bollen, K. A. (1989). Structural equations with latent variables. New York: Wiley.

- Buchwald, J. Z. (2006). Discrepant measurements and experimental knowledge in the early modern era. Archive for History of Exact Sciences, 60, 565–649.

- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81–105.

- Carmines, E. G., & McIver, J. P. (1981). Analyzing models with unobserved variables: Analysis of covariance structures. In G. W. Bohrnstedt & B. F. Borgatta (Eds.), Social measurement:Current issues (pp. 65–115). Beverly Hills, CA: Sage.

- Cattell, R. B. (1966). The screen test for the number of factors. Multivariate Behavioral Research, 1, 245–276.

- Choi, S. W., Reise, S. P., Pilkonis, P. A., Hays, R. D., & Cella, D. (2010). Efficacy of static and computer adaptive short forms compared to full-length measure of depressive symptoms. Quality of Life Research, 19, 125–136.

- Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37–46.
- Cohen, P., Cohen, J., Teresi, J., Marchi, M., & Velez, C. N. (1990). Problems in the measurement of latent variables in structural equation causal models. Applied Psychological Measurement, 14, 183–196.
- Comrey, A. L. (1973). A first course in factor analysis. New York: Academic Press.
- Comrey, A. L. (1988). Factor analytic methods of scale development in personality and clinical psychology. Journal of Consulting and Clinical Psychology, 56, 754–761.
- Converse, J. M., & Presser, S. (1986). Survey questions: Handcrafting the standardized questionnaire. Sage university paper series on quantitative applications in the social sciences (Series No. 07-063). Beverly Hills, CA: Sage.
- Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart & Winston.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297–334.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). Dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: Wiley.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52, 281–302.
- Cureton, E. E. (1983). Factor analysis: An applied approach. Hillsdale, NJ: Lawrence Erlbaum.

- Currey, S. S., Callahan, L. F., & DeVellis, R. F. (2002). Five-item Rheumatology Attitudes Index (RAI): Disadvantages of a single positively worded item. Unpublished paper, Thurston Arthritis Research Center, University of North Carolina at Chapel Hill.
- Czaja, R., & Blair, J. (1996). Designing surveys: A guide to decisions and procedures. Thousand Oaks, CA: Pine Forge.
- Dale, F., & Chall, J. E. (1948). A formula for predicting readability: Instructions. Education Research Bulletin, 27, 37–54.
- Davey, Z. Z. T., & Hendrickson, A. (2010, May). Classical versus IRT statistical test specifications for building test forms. Paper presented at the annual meeting of the National Council of Measurement Education, Denver, Colorado.
- De Boeck, P., & Wilson, M. (2004). Explanatory item response models: A generalized linear and nonlinear approach. New York: Springer.
- Decker, H. S. (2007). How Kraepelinian was Kraepelin? How Kraepelinian are the neo-Kraepelinians? From Emil Kraepelin to DSM-III. History of Psychiatry, 18(3), 337–360.
- DeVellis, R. F. (1996). A consumer's guide to finding, evaluating, and reporting on measurement instruments. Arthritis Care and Research, 9, 239–245.
- DeVellis, R. F. (2005). Inter-rater reliability. In K. Kempf-Leonard (Ed.), Encyclopedia of social measurement (Vol. 2, pp. 317–322). San Diego: Elsevier.
- DeVellis, R. F., & Callahan, L. F. (1993). A brief measure of helplessness: The helplessness subscale of the Rheumatology Attitudes Index. Journal of Rheumatology, 20, 866–869.

- DeVellis, R. F., DeVellis, B. M., Blanchard, L. W., Klotz, M. L., Luchok, K., & Voyce, C. (1993). Development and validation of the Parent Health Locus of Control (PHLOC) scales. Health Education Quarterly, 20, 211–225.
- DeVellis, R. F., DeVellis, B. M., Revicki, D. A., Lurie, S. J., Runyan, D. K., & Bristol, M. M. (1985). Development and validation of the child improvement locus of control (CILC) scales. Journal of Social and Clinical Psychology, 3, 307–324.
- DeVellis, R. F., Holt, K., Renner, B. R., Blalock, S. J., Blanchard, L. W., Cook, H. L., et al. (1990). The relationship of social comparison to rheumatoid arthritis symptoms and affect. Basic and Applied Social Psychology, 11, 1–18.
- DeWalt, D. A., Rothrock, N., Yount, S., & Stone, A. A. (2007). Evaluation of item candidates: The PROMIS qualitative item review. Medical Care, 45(5, Suppl. 1), S12–S21.
- Dillman, D. A. (2007). Mail and Internet surveys: The tailored design (2nd ed., 2007 update). Hoboken, NJ: Wiley.
- Duncan, O. D. (1984). Notes on social measurement: Historical and critical. New York: Russell Sage.
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. British Journal of Psychology, 105, 399–412.
- Embretson, S. E., & Hershberger, S. L. (1999). Summary and future of psychometric models in testing. In S. E. Embretson & S. L. Hershberger (Eds.), The new rules of measurement (pp. 243–254).

- Mahwah, NJ: Lawrence Erlbaum. Embretson, S. E., & Reise, S. P. (2010). Item response theory (2nd ed.). New York: Routledge Academic.
- Eysenck, H. J. (1968). Kraepelin, Emil. International Encyclopedia of the Social Sciences. Retrieved October 9, 2015, from http://www. encyclopedia. com/doc/1G2-3045000662.html.
- Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. Educational and Psychological Measurement, 58, 357–381.
- Festinger, L. (1954). A theory of social comparison processes. Human Relations, 7, 117–140.
- Fink, A. (1995). The survey kit. Thousand Oaks, CA: Sage.
- Fowler, F. J., Jr. (2009). Survey research methods (4th ed.). Thousand Oaks, CA: Sage.
- Fry, E. (1977). Fry's readability graph: Clarifications, validity, and extension to level 17. Journal of Reading, 21, 249.
- Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. Practical Assessment, Research and Evaluation, 17, 1–13.
- Ghiselli, B. E., Campbell, J. P., & Zedeck, S. (1981). Measurement theory for the behavioral sciences. San Francisco: Freeman.
- Gorsuch, R. L. (1983). Factor analysis. Hillsdale, NJ: Lawrence Erlbaum.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. Newbury Park, CA: Sage.
- Harman, H. H. (1976). Modern factor analysis. Chicago: University of Chicago Press.

- Hathaway, S. R., & McKinley, J. C. (1967). Minnesota Multiphasic Personality Inventory: Manual for administration and scoring. New York: Psychological Corporation.

- Hathaway, S. R., & Meehl, P. E. (1951). An atlas for the clinical use of the MMPI. Minneapolis: University of Minnesota Press.

- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. Organizational Research Methods, 7(2), 191–205.

- Hewett, J. E., Anderson, S. K., & Minor, M. A. (1992). Statistical analyses for research in arthritis when assessing change. Arthritis Care and Research, 5, 74–80.

- Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. Journal of Health and Social Behavior, 38, 21–37.

- Insel, T. R. (2013). Director's blog: Transforming diagnosis. Accessed July 21, 2015, from http://www.nimh.nih.gov/about/director/2013/transforming-diagnosis.shtml.

- Insel, T. R. (2015). Dr. Tom Insel to step down as NIMH director. Accessed October 9, 2015, from http://www.nimh.nih.gov/about/drtom- insel-to-step-down-as-nimh-director.shtml.

- Insel, T. R., & Leiberman, J. A. (2013). DSM-5 and RDoC: Shared interests. Retrieved September 18, 2015, from www.nimh.nih.gov/news/ science-news/2013/dsm-5-and-rdoc-shared-interests.shtml.

- Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. Psychometrika, 36, 109–134.

- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 141–151.
- Keefe, F. J. (2000). Self-report of pain: Issues and opportunities. In A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), The science of self-report: Implications for research and practice (pp. 317–337). Mahwah, NJ: Lawrence Erlbaum.
- Kelly, J. R., & McGrath, J. B. (1988). On time and method. Newbury Park, CA: Sage.
- Kihlstrom, J. F., Eich, E., Sandbrand, D., & Tobias, B. A. (2000). Emotion and memory: Implications for self-report. In A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), The science of self-report: Implications for research and practice (pp. 81–103). Mahwah, NJ: Lawrence Erlbaum.
- Kirk, R. E. (1995). Experimental design: Procedures for the behavioral sciences (3rd ed.). San Francisco: Brooks/Cole.
- Kraepelin, E. (1883). Compendium der Psychiatrie zum Gebrauche für Studirende und Aerzte. Leipzig: Abel.
- Krueger, R. A., & Casey, M. A. (2000). Focus groups: A practical guide for applied research. Thousand Oaks, CA: Sage.
- Kreuger, R. F., Watson, D., & Barlow, D. H. (2005). Introduction to the special section: Toward a dimensionally based taxonomy of psychopathology. Journal of Abnormal Psychology, 114, 491–493.
- Kruyen, P. M., Emons, W. H. M., & Sijtsma, K. (2014). Assessing individual change using short tests and questionnaires. Applied Psychological Measurement, 38, 201–216.

- Levenson, H. (1973). Multidimensional locus of control in psychiatric patients. Journal of Consulting and Clinical Psychology, 41, 397–404.
- Linn, R. L., & Slinde, J. A. (1977). The determination of the significance of change between pre- and post-testing periods. Review of Educational Research, 47, 121–150.
- Lipsey, M. W. (1990). Design sensitivity: Statistical power for experimental research. Newbury Park, CA: Sage.
- Loehlin, J. C. (1998). Latent variable models: An introduction to factor, path, and structural analysis. Mahwah, NJ: Lawrence Erlbaum.
- Long, J. S. (1983). Confirmatory factor analysis. Sage university paper series on quantitative applications in the social sciences (Series No. 07-033). Beverly Hills, CA: Sage.
- Lord, F. M., & Novick, M. R. (2008). Statistical theories of mental test scores. Charlotte, NC: Information Age.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. Psychological Methods, 4, 84–99.
- Mayer, J. M. (1978). Assessment of depression. In M. P. McReynolds (Ed.), Advances in psychological assessment (Vol. 4, pp. 358–425). San Francisco: Jossey-Bass.
- McDonald, R. P. (1984). Factor analysis and related methods. Hillsdale, NJ: Lawrence Erlbaum.
- Meehl, P. E. (1999). Clarifications about taxometric method. Applied & Preventive Psychology, 8, 165–174.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50, 741–749.

- Mitchell, S. K. (1979). Interobserver agreement, reliability, and generalizability of data collected in observational studies. Psychological Bulletin, 86, 376–390.

- Mlodinow, L. (2008). The drunkard's walk: How randomness rules our lives. New York: Pantheon.

- Murphy, L. L., Spies, R. A., & Plake, B. S. (2006). Tests in print. Lincoln: University of Nebraska Press.

- Myers, J. L. (1979). Fundamentals of experimental design (3rd ed.). Boston: Allyn & Bacon.

- Namboodiri, K. (1984). Matrix algebra: An introduction. Sage university paper series on quantitative applications in the social sciences (Series No. 07-028). Beverly Hills, CA: Sage.

- Narens, L., & Luce, R. D. (1986). Measurement: The theory of numerical assignments. Psychological Bulletin, 99, 166–180.

- Nering, M. L., & Ostini, R. (2010). Handbook of polytomous item response theory models. New York: Routledge.

- Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.

- Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). New York: McGraw-Hill.

- Osgood, C. E., & Tannenbaum, P. H. (1955). The principle of congruence in the prediction of attitude change. Psychological Bulletin, 62, 42–55.

- Overall, J. E., & Woodward, J. A. (1975). Unreliability of difference scores: A paradox for measurement of change. Psychological Bulletin, 82(1), 85–86.

- Padilla, M. A., Divers, J., & Newton, M. (2013). Coefficient alpha bootstrap confidence interval under normality. Applied Psychological Measurement, 36, 331–348.
- PBS. (2002). Timeline: Treatments for mental illness. Retrieved July 21, 2015, from http://www.pbs.org/wgbh/amex/nash/timeline.
- Peters, G. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. The European Health Psychologist, 16, 56–69.
- PROMIS. (2010). Patient-Reported Outcomes Measurement Information System: Dynamic tools to measure health outcomes from the patient perspective.
- Retrieved from http://www.nihpromis.org/default.aspx
- Radloff, L. (1977). The CES-D scale: A self-report depression scale for research in the general population. Applied Psychological Measurement, 1, 385–401.
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Chicago: MESA.
- Regalado, A. (2015, September 21). Why America's top mental health researcher joined Alphabet. Retrieved October 15, 2015, from http://www.technologyreview.com/news/541446/why-americas-topmental-health-researcher-joined-alphabet.
- Reise, S. P., Moore T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. Journal of Personality Assessment, 92, 544–559.

- Reise, S. P., & Revicki, D. A. (2015). Handbook of item response theory modeling: Applications to typical performance assessment. New York: Routledge.
- Rholes, W. S., Riskind, J. H., & Lane, J. W. (1987). Emotional states and memory biases: Effects of cognitive priming and mood. Journal of Personality and Social Psychology, 52, 91–99.
- Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). Measures of personality and social psychological attitudes. San Diego: Academic Press.
- Rogosa, D. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. Meredith, & S. C. Rawlings (Eds.), Methodological issues in aging research (pp. 171–210). New York: Springer.
- Rosenberg, M. (1965). Society and the adolescent self-image. Princeton, NJ: Princeton University Press.
- Rotter, J. B. (1966). Generalized expectancies for internal vs. external control of reinforcement. Psychological Monographs, 80(1, Whole No. 609).
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. Psychometric Monograph (Suppl. 17).
- Saucier, G., & Goldberg, L. R. (1996). The language of personality: Lexical perspectives on the five-factor model. In J. S. Wiggins (Ed.), The five-factor model of personality (pp. 21–50). New York: Guilford.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86, 420–428.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. Psychometrika, 74, 107–120.

- Silvestro-Tipay, J. L. (2009). Item response theory and classical test theory: An empirical comparison of item/person statistics in a biological science test. International Journal of Educational and Psychological Assessment, 1, 19–31.
- Smith, P. H., Earp, J. A., & DeVellis, R. F. (1995). Measuring battering: Development of the Women's Experiences with Battering (WEB) scale. Women's Health: Research on Gender, Behavior, and Policy, 1, 273–288.
- Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). State-trait anxiety inventory (STAI) test manual for form X. Palo Alto, CA: Consulting Psychologists Press.
- Spies, R. A., Carlson, J. F., & Geisinger, K. F. (2010). The eighteenth mental measurements yearbook. Lincoln: University of Nebraska Press.
- Stage, C. (2003). Classical test theory or item response theory: The Swedish experience (EM No. 42). Umea, Sweden: Umea Universitet Department of Educational Measurement. Available online at http:// www8.umu.se/edmeas/publikationer/index_eng.html
- Sterba, K. R., DeVellis, R. F., Lewis, M. A., Baucom, D. H., Jordan, J. M., & DeVellis, B. M. (2007). Developing and testing a measure of dyadic efficacy for married women with rheumatoid arthritis and their spouses. Arthritis & Rheumatism (Arthritis Care & Research), 57(2), 294–302.
- Strahan, R., & Gerbasi, K. (1972). Short, homogenous version of the Marlowe-Crowne Social Desirability Scale. Journal of Clinical Psychology, 28, 191–193.
- Tartakovsky, M. (2011). How the DSM developed: What you might not know. Retrieved October 9, 2015, from http://psychcentral. com/ blog/archives/2011/07/02/how-the-dsm-developed-whatyou- might-not-know.

- Thissen, D., Chen, W.-H., & Bock, R. D. (2003). Multilog (Version 7.0) [Computer software]. Lincolnwood, IL: Scientific Software International.
- Timmermann, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. Psychological Methods, 16, 209–220.
- Tinsley, H. E. A., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. Journal of Counseling Psychology, 34, 414–424.
- Traub, R. E. (1994). Reliability for the social sciences: Theory and applications.
- Thousand Oaks, CA: Sage.
- Van der Linden, W. J., & Glas, C. A. W. (2000). Computerized adaptive testing: Theory and practice. St. Paul, MN: Assessment Systems.
- Venables, W. N., Smith, D. M., & the R-Core team. (2015). An introduction to R. Notes on R: A programming environment for data analysis and graphics. Available for download from https://cran.r-project.org/ manuals. html.
- Wallston, K. A., Stein, M. J., & Smith, C. A. (1994). Form C of the MHLC Scales: A condition-specific measure of locus of control. Journal of Personality Assessment, 63, 534–553.
- Wallston, K. A., Wallston, B. S., & DeVellis, R. (1978). Development and validation of the multidimensional health locus of control (MHLC) scales. Health Education Monographs, 6, 161–170.
- Weisberg, H., Krosnick, J. A., & Bowen, B. D. (1996). An introduction to survey research, polling, and data analysis. Thousand Oaks, CA: Sage.

- Willis, G. B. (2005). Cognitive interviewing: A tool for improving questionnaire design. Thousand Oaks, CA: Sage.
- Wright, B. D. (1999). Fundamental measurement for psychology. In S. E. Embretson & S. L. Hershberger (Eds.), The new rules of measurement (pp. 65–104). Mahwah, NJ: Lawrence Erlbaum.
- Yu, C. H. (2005). Test-retest reliability. In K. Kempf-Leonard (Ed.), Encyclopedia of social measurement (Vol. 3, pp. 777–784). San Diego, CA: Elsevier.
- Zickar, M. J., & Broadfoot, A. A. (2008). The partial revival of a dead horse? Comparing classical test theory and item response theory. In C. E. Lance & R. J. Vandenberg (Eds.), Statistical and methodological myths and urban legends. New York: Routledge Academic.
- Zorzi, M., Priftis, K., & Umilita, C. (2002, May). Brain damage: Neglect disrupts the mental number line. Nature, 417, 138–139.
- Zuckerman, M. (1983). The distinction between trait and state scales is not arbitrary: Comment on Allen and Potkay's "On the arbitrary distinction between traits and states." Journal of Personality and Social Psychology, 44, 1083–1086.
- Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. Psychological Bulletin, 99, 432–442.

المترجمان في سطور

١- د. سعد بن سعيد القحطاني

المؤهلات العلمية:

حاصـل على درجة الدكتـوراه في الإحصاء التطبيقي عام ٢٠١٠م مـن جامعة كولورادو بالولايات المتحدة الأمريكية.

الوظيفة الحالية:

مدير مشروع قياس الإنتاجية وأمين المجلس العلمي بمعهد الإدارة العامة.

أهم الخبرات العلمية والعملية:

- مدير سابق لقطاع الإحصاء وإدارة الدعم المنهجي والإحصائي بالمعهد.
- مستشار إحصائي غير متفرغ لعدد من الأجهزة الحكومية.
- مستشار إحصائي بمركز الاستشارات والبحوث بجامعة كلورادو الشمالية بالولايات المتحدة الأمريكية لمدة ثلاث سنوات.
- تحكيم عدد كبير من البحوث والدراسات العلمية والاستبانات.
- شارك في رئاسة وعضوية العديد من اللجان العلمية والاستشارية داخل وخارج المعهد.
- تقديـم العديد من المحـاضرات والدورات التدريبية والحلقـات التطبيقية وورش العمل واللقـاءات العلمية والمشـاركات الإعلامية داخل وخارج معهـد الإدارة العامة في مجال الإحصاء وتحليل البيانات وتصميم الاستبانات ومؤشرات الأداء.
- تأليف وترجمة عدد من الكتب والحقائب التدريبية والتي من أهمها كتاب مؤلف في الإحصاء التطبيقي (٢٠١٥م) وآخر مترجم في نمذجة المعادلات البنائية (SEM)باستخدام أموس.
- نال جائزة عميد الدراسات العليا للتفوق الأكاديمي في مرحلة الدكتوراه، الميدالية الذهبية كجائـزة لأفضل أطروحة دكتوراه على مسـتوى الجامعة، جائـزة أفضل مدرب في معهد الإدارة العامة لعام ٢٠١٦م.

٢- أ.د. طارق عطية عبدالرحمن

المؤهلات العلمية:

دكتوراه في علم الاجتماع الريفي من جامعة كفر الشيخ- مصر، ٢٠٠٧م.

الوظيفة الحالية:

أســتاذ دكتور في علم الاجتماع الريفي بجامعة كفر الشــيخ، مصــر؛ وعضو هيئة تدريب بمركز البحوث والدراسات بمعهد الإدارة العامة، الرياض.

أهم الخبرات العلمية والعملية:

- كتابــة ونشر أكثر من (٤٢) عملاً علمياً نُشرت في مجــلات علمية محكمة عربية، وعالمية (indexed scoups journals).
- تأليف عدد من الكتب العلمية المحكمة في مجالات علم الاجتماع الريفي، ورأس المال الاجتماعي والفعالية التنظيمية، وإدارة المعرفة والإبداع التنظيمي، وأساليب ومناهج البحث العلمي.
- ترجمة كتاب في أساسيات اختيار أساليب المعاينة.
- تقديم العديد من المحاضرات والدورات التدريبية والحلقات التطبيقية وورش العمل واللقاءات العلمية والاستشارات في مجال منهجية البحث العلمي والإحصاء الاجتماعي وتحليل البيانات.
- تأليف وتصميم ومراجعة عدد من الحقائب التدريبية والبرامج التدريبية والملفات العلمية.
- المشاركة في عدد من المشروعات البحثية كباحث رئيس (PI) أو باحث مشارك (COI) في عدد من الجهات البحثية.
- عمل عضواً في لجنة البحوث العلمية بمعهد الإدارة العامة بالرياض.
- عمل عضواً في هيئة تحرير مجلة "الإدارة العامة".
- محكِّم معتمد لدى عدد من الجهات البحثية والمجلات العلمية.
- الإشراف على عدد من طلاب الدراســات العليا لدرجتَي الماجستير والدكتوراه بجامعة كفر الشيخ، مصر.

مراجع الترجمة في سطور

أ.د. عبدالرحمن بن سليمان العنقري

المؤهلات العلمية:

دكتـوراه في الإحصـاء التطبيقي، تخصـص إحصاء لامعلمي، جامعة جلاسـجو، المملكة المتحدة، ١٩٩٥.

الوظيفة الحالية:

مستشار متفرغ ومشرف عام على مركز إحصاءات التعليم العالي، وزارة التعليم العالي.

أهم الخبرات العلمية والعملية:

- مستشار متفرغ لدى وزارة التعليم العالي من عام ١٤٢٦هـ حتى الآن.
- مستشار إحصائي لفريق الخطة الوطنية لتقنية المعلومات من عام ١٤٢٤ حتى ١٤٣٤هـ.
- باحث زائر ، كلية العلوم، الجامعة الإسلامية في ماليزيا، ماليزيا، ١٤٢٣-١٤٢٤هـ.
- رئيس قسم الإحصاء وبحوث العمليات، كلية العلوم، جامعة الملك سعود، ١٤١٨- ١٤٢٢هـ.
- رئيس الجمعية السعودية للعلوم الرياضية، جامعة الملك سعود، ١٤٢٥-١٤٢٩هـ.
- متعاون مع معهد الإدارة العامة بالرياض لتقديم دورات في طرق اختيار العينات وتصميم الاستبانة الإحصائية، ١٤١٧هـ.
- متعاون مع المؤسسة العامة للتدريب التقني والمهني لتدريس مقرر الاحتمالات والإحصاء الهندسي بالكلية التقنية بالرياض، ١٤١٩هـ.
- مناقشـة عدد من رسـائل الدكتوراه والماجستير بجامعة الملك سعود وجامعة الأميرة نورة بنت عبد الرحمن.
- الإشراف المباشر على نظام إحصاءات التعليم العالي الإلكتروني.
- الإشراف المباشر على صفحة مركز إحصاءات التعليم العالي على موقع وزارة التعليم العالي، ومتابعة تحديثها.

تم التصميم والإخراج الفني والطباعة في
الإدارة العامة للطباعة والنشر بمعهد الإدارة العامة - ١٤٤٢هـ

هذا الكتاب:

يأتي هذا الكتاب ليقدِّم لنا حصيلةَ وخلاصة خبرة مؤلفه الدكتور/ ديفيليس التي تزيد عن (٣٥) عاماً في مجال القياس. حيث يهدف الكتابُ بصفةٍ رئيسة إلى تسليط الضوء على المفاهيم والنظريات والاتجاهات التقليدية والحديثة المتعلقة بتطوير المقاييس في مجال البحث العلمي. ويُعدُّ من أهم الكتب في مجال تطوير المقاييس ومن أكثرها مبيعاً في نسخته الأولى وفقاً لموقع Google Books.

ويتمتع الكتابُ بالعديد من المزايا أبرزها: إزالته للغموض واللبس الذي يعتري عملية القياس العلمي وبناء المقاييس في البحوث الاجتماعية والعلوم الأخرى ذات الصلة، من خلال التأكيد على الفهم المنطقي الدقيق للمفاهيم المرتبطة ببناء وتطوير المقاييس. كما يساعد الكتاب القراء على فهم الأساليب الحديثة للقياس وبناء المقاييس، ومقارنتها بالأساليب والطرق الكلاسيكية، والفهم العميق للمزايا النسبية لكلٍ منها. هذا عوضاً عن لغته السهلة الواضحة البعيدة عن الحسابات المعقدة وبالتالي فهو مصدرٌ للباحثين ذوي الخبرة والباحثين المبتدئين والطلاب على حدٍ سواء.

ويتناول هذا الكتاب في طبعته الرابعة، المكونة من ثمانية فصول، موضوعاتٍ جديدة ذات صلة بمناهج القياس الحديثة، منها على سبيل المثال بدائل معامل "ألفا"، والنموذج ذي العاملين، ونموذج العوامل الهرمي، والأدوات المتاحة في البرامج الإحصائية ذات العلاقة بتطوير المقاييس في البحوث العلمية. كما تضمَّن أمثلةً عملية وتمارين للمساعدة في التعلم وفهم أساسيات القياس.



تصميم وإخراج وطباعة الإدارة العامة للطباعة والنشر ـ معهد الإدارة العامة ١٤٤٢هـ